UNIVERSITÉ LAVAL

Prêts pair-à-pair : Analyse des multi-prêteurs, du langage et de la formation de groupes

Mémoire

Bérénice Agbodossindji

Maîtrise en sciences de l'administration - avec mémoire
Maître ès sciences (M. Sc.)

Québec, Canada

© Bérénice Agbodossindji, 2022

Prêts pair-à-pair: Analyse des multi-prêteurs, du langage et de la formation de groupes

Mémoire

Bérénice Agbodossindji

Sous la direction de :

Philippe Grégoire, directeur de recherche
Richard Luger, codirecteur de recherche

Résumé

Cette étude vise à mieux comprendre les choix d'Investissement d'un groupe particulier de crowdfunders, en l'occurrence ceux qui financement un grand nombre de projets (n ⩾10). Pour ce faire, nous utilisons les données de Kiva, la plateforme en ligne la plus importante en matière de financement participatif à but non lucratif, et désignons les prêteurs concernés sous le terme de « multi-prêteurs ». Premièrement, nous évaluons le niveau de diversification des portefeuilles de prêts des multi-prêteurs et examinons comment leur capacité à sélectionner les meilleurs projets varie en fonction du niveau de diversification adopté. Deuxièmement, nous étudions le lien entre le langage intrinsèque (langage décrivant un prêt comme une opportunité d'aider les personnes dans le besoin) et la durée des campagnes d'une part et le lien entre le langage extrinsèque (langage décrivant un prêt comme une opportunité rentable) et la durée des campagnes d'autre part. Enfin, nous examinons l'effet d'être un groupe d'emprunteurs plutôt qu'un emprunteur seul sur la durée des campagnes de financement.

Les résultats obtenus révèlent qu'il y a trois niveaux de diversification (faible, moyen et élevé) et que la majorité des multi-prêteurs adoptent une diversification moyenne. Ils montrent également qu'une diversification moyenne permet aux multi-prêteurs de choisir des projets qui réussissent leur campagne de collecte de fonds en peu de temps. En ce qui concerne le langage intrinsèque et le langage extrinsèque, nous obtenons des résultats qui vont à l'encontre de ce qui avait été préalablement remarqué dans les débuts de la plateforme Kiva. En effet, le langage intrinsèque s'est révélé être associé à une augmentation de la durée des campagnes de financement et le langage extrinsèque à une réduction de cette durée. Quant aux projets qui n'ont pas bénéficié de la contribution des multi-prêteurs, les résultats indiquent que le langage intrinsèque et le langage extrinsèque sont sans incident sur la durée des campagnes de financement. Par ailleurs, cette étude met en évidence un changement dans le comportement de prêts des prêteurs Kiva au fil du temps. En effet, nous avons remarqué, qu'avant 2010 le langage intrinsèque était associé à une réduction du temps nécessaire au financement complet et que le langage extrinsèque était associé à une augmentation de ce temps mais qu'à partir de 2010, cela s'est inversé. Enfin, pour le troisième objectif nous montrons qu'être un groupe d'emprunteurs plutôt qu'un emprunteur seul réduit le temps nécessaire jusqu'au financement complet des projets.

Mots clés : financement participatif, multi-prêteurs, diversification, langage, prêts de groupe, Kiva.

Table des matières

Résumé ii
Table des matières ii
Liste des tableaux v
Liste des figures vi
Liste des abréviations, sigles, acronymes vii
Remerciements ix
Introduction 1
1. Bref aperçu du marché de prêts pair-à-pair 4
1.1. Du crowdsourcing au crowdfunding 4
1.2. Le financement participatif basé sur la dette 6
1.3. Les motivations à la participation au marché de prêts pair-à-pair 9
1.4. Les risques et défis du marché des prêts pair-à-pair 11
2. Revue de littérature 14
2.1. Revue de littérature sur la diversification des portefeuilles de prêts 14
2.2. Revue de littérature sur les motivations intrinsèque et extrinsèque 15
2.3. Revue de littérature sur les récits de demande de prêt 17
2.4. Revue de littérature sur les prêts de groupe 20
2.5. Conclusion de la revue de littérature 22
3. Objectifs et hypothèses de recherche 23
4. Méthodologie 25
4.1. Diversification des portefeuilles de prêts 25
4.1.1. Mesures de diversification des portefeuilles 25
4.1.2. Construction d'un indicateur synthétique 27
4.1.3. Méthodes de classification des portefeuilles 30
4.1.4. Régression des risques proportionnels de Cox 31
4.2. Récits de demande de prêt et durée des campagnes de financement 35
4.3. Prêts de groupe versus prêts individuels 37
4.3.1. Le cadre des résultats potentiels 37

4.3.2. Inférence causale en présence de données non randomisées de grande dimension 40
4.3.3. Méthodes d'estimation de l'effet causal 42
4.3.4. Modèles de régression des résultats et du score de propension 48
5. Description des données et mesures des variables 58
5.1. Données d'étude 58
5.2. Définition et mesure des variables 61
5.2.1. Variables dépendantes 61
5.2.2. Variables indépendantes 62
5.2.3. Variables de contrôle 63
6. Analyse descriptive 67
7. Résultats et discussion 81
7.1. Principaux résultats sur la diversification des portefeuilles de prêts 81
7.1.1. Résultats des mesures de diversifications et construction de l'indicateur 81
7.1.2. Résultats de la classification 85
7.1.3. Résultats de la performance des classes 88
7.2. Principaux résultats sur les récits de demande de prêts 92
7.2.1. Résultats sur l'intérêt à étudier les récits de demandes de prêts des multi-prêteurs 92
7.2.2. Résultats sur le changement dans le comportement de prêts des prêteurs Kiva 95
7.2.3. Résultats du langage intrinsèque et extrinsèque pour l'échantillon des multi-prêteurs 98
7.2.4. Mesure de la sensibilité des résultats MCO 101
7.3. Principaux résultats sur les prêts de groupe 103
7.3.1. Analyse des résultats de métriques d'évaluation 103
7.3.2. Analyse des résultats de l'effet moyen de traitement 105
Conclusion 109
Bibliographie 112
Annexe A : Description des variables 126
Annexe B : Exemples de récits 130
Annexe C : Liste des pays des emprunteurs et nombre de demandes de prêts par pays 131

Annexe D : Coefficients de corrélation 133
Annexe E : Résultats de MCO pour l'échantillon Éch3b 134
Annexe F : Résultats de l'algorithme génétique 135

Liste des tableaux

Tableau 1 : Nombre de multi-prêteurs par pays, nombre de contributions par pays et per capita 67
Tableau 2 : Sommaire statistique sur la composition des portefeuilles de prêts 69
Tableau 3 : Sommaire de statistiques sur les variables quantitatives 71
Tableau 4 : Comparaison des échantillons Éch2b et Éch3b 72
Tableau 5 : Nombre de prêts par groupe et individu, genre et risque 73
Tableau 6 : Sommaire statistique de la durée et des montants en fonction du groupe, genre et risque 75
Tableau 7 : Résultats de la régression du montant de prêts sur la durée des campagnes 76
Tableau 8 : Nombre de prêts par groupe, genre et risque en fonction des secteurs d'activité (Fréquence relative en %) 77
Tableau 9 : Moyennes et Écart-types de la durée et des montants par secteur d'activité 79
Tableau 10 : Sommaire des mesures de diversification 82
Tableau 11 : Indices MSA et KMO 82
Tableau 12 : Valeur propre, variance expliquée et cumulée par axe 82
Tableau 13 : Coordonnées des variables par axe 84
Tableau 14 : Statistiques de l'indicateur normalisé et non normalisé 84
Tableau 15 : Résultats de classification 87
Tableau 16 : Niveau de diversification par classe 88
Tableau 17 : Durée moyenne et % de campagnes réussies (classe 1 et classe 2) 88
Tableau 18 : Durée moyenne et % de campagnes réussies (classe 1 et classe 3) 89
Tableau 19 : Durée moyenne et % de campagnes réussies (classe 2 et classe 3) 89
Tableau 20 : Résultats de la régression de Cox 91
Tableau 21 : Résultats de la régression MCO avec les effets d'interaction 94
Tableau 22 : Résultats de la régression MCO de 2006 à 2009 et de 2010 à 2019 97
Tableau 23 : Résultats de MCO pour l'échantillon des multi-prêteurs (Éch2b) 100
Tableau 24 : Récapitulatif des résultats MCO et COX 102
Tableau 25 : Résultats des métriques d'évaluation 103
Tableau 26 : Résultats des méthodes 105

Liste des figures

Figure 1 : Étapes de l'algorithme génétique 57
Figure 2 : Composition des portefeuilles de prêts 70
Figure 3 : Pourcentage de prêts par secteur d'activité 76
Figure 4 : Représentation des valeurs propres 83
Figure 5: Dendrogramme 85
Figure 6 : Résultats des 30 indices (NbClust) 86

Liste des abréviations, sigles, acronymes

ACP : Analyse en composantes principales

ECMV : Estimateur ciblé du maximum de vraisemblance

EDR : Estimateur doublement robuste

GloVe : Global Vectors

IMF : Institution de microfinance

LMCT : Longue mémoire à court terme

P2P : Pair-à-Pair ou Particulier à Particulier

À celui qui est, qui était et qui vient…

Remerciements

Mes remerciements les plus sincères vont à mon directeur de recherche, le professeur Philippe Grégoire. Sans son appui scientifique et ses conseils avisés tout au long du processus, ce mémoire n'aurait pas pu voir le jour. Je suis très reconnaissante pour sa patience, sa générosité débordante et sa qualité humaine de compréhension.

Je tiens également à remercier mon codirecteur de recherche, le professeur Richard Luger, qui a été présent à mes débuts à la FSA. Il a toujours été disponible pour répondre à toutes mes questions économétriques. Ses commentaires d'expert ont permis d'enrichir et d'améliorer la qualité de ce travail.

À ma famille, je dis un grand merci pour son soutien à tous égards. Je reste convaincue d'avoir vu le jour dans la meilleure famille au monde.

Introduction

Initiée dans les années 70 au Bangladesh, grâce à la Grameen Bank, la microfinance est apparue comme une arme efficace pour la lutte contre la pauvreté dans les pays du tiers monde. Elle consiste essentiellement à accorder de très petits prêts à des entrepreneurs démunis en vue d'augmenter l'activité entrepreneuriale et par là même stimuler la croissance des pays concernés (Armendáriz de Aghion et Morduch, 2005; Yunus, 2007 ; Duvendack et al., 2011). Ainsi, les demandeurs de prêt qui autrefois étaient exclus des services financiers classiques parce qu'ils ne disposaient pas de garanties ou d'antécédents de crédit pouvaient accéder au capital financier recherché et ce à des taux d'intérêt bas (Armendáriz de Aghion et Morduch, 2005 ; Khavul, 2010). Au fil des années, la microfinance a gagné du terrain conduisant à la création de multiples et diverses institutions de microfinance (IMF) de par le monde parmi lesquelles on trouve des fonds institutionnels, des fondations ainsi que des investisseurs qui utilisent des plateformes en ligne (Khavul, 2010).

Les plateformes en ligne, en tant qu'intermédiaires dans le financement participatif basé sur internet, ont donné un nouvel essor à la microfinance. En effet, elles offrent aux IMF un canal adéquat pour collecter les petits prêts d'innombrables prêteurs du monde entier afin de répondre aux besoins de financement des emprunteurs. En 2014, le volume de financement effectué sur ces plateformes s'est chiffré à 16,2 milliards de dollars, représentant 166 % de celui en 2013 (Massolution, 2017)[1]. Avec la pandémie de la COVID-19, le volume de 2020 est de 17,2 milliards de dollars et est destiné à atteindre 34,6 milliards de dollars d'ici 2026 (Market Data Forecast, 2021)[2].

Pour le segment le plus prospère de ce marché, le crowdlending encore appelé le prêt pair-à-pair, la crise mondiale du crédit a été plus que bénéfique. Par exemple, entre avril 2008 et septembre 2008, le volume de prêts financés sur Prosper a augmenté de 41 % par rapport aux six mois précédents et depuis cette crise, le prêt P2P est devenu une option alternative attrayante pour les emprunteurs qui ont peu de chances d'obtenir un prêt auprès d'une banque traditionnelle. Sur Kiva, environ 2,5 millions de dollars de prêts sont financés en moyenne chaque semaine.

[1] http://crowdexpert.com/crowdfunding-industry-statistics/

[2] https://www.marketdataforecast.com/market-reports/crowdfunding-market

Cependant, il existe un revers pas très commode de la médaille car dans ce type de marché toutes les interactions entre le prêteur et l'emprunteur se font par l'intermédiaire de l'interface du site des plateformes. Cette forme d'interaction rend les investissements en financement participatif plus risqués que les investissements traditionnels (Yum et al., 2012). Deux questions nous viennent donc à l'esprit : Pourquoi les investisseurs choisissent-ils de participer au financement participatif? Comment choisissent-ils les emprunteurs à qui ils accordent leur soutien? Les recherches antérieures ont répondu à la première question en se basant sur les motivations des investisseurs et ont montré qu'elles sont pour certains investisseurs intrinsèques (Gerber et Hui, 2013; Bretschneider et Leimeister, 2017; McKinnon et al. 2013; Liu et al. 2012), mais pour d'autres extrinsèques (Cumming et Johan, 2013; Cholakova et Clarysse, 2015; Collins et Pierrakis, 2012). Pour la seconde question, les études antérieures ont identifié plusieurs facteurs qui font que certaines campagnes de financement réussissent mieux que d'autres. Nous avons, par exemple, le contenu des récits de demande de prêts et la manière dont ce contenu est libellé (Larrimore et al., 2011; Herzenstein et al., 2011; Kim et al., 2016; Parhankangas et Renko, 2017; Moss et al., 2017). Également, le fait que l'emprunteur soit un groupe ou un individu ainsi que son capital social (Ly et Mason, 2012b; Galak et al., 2011; Zhang et Liu, 2012; Burtch et al., 2013; Kim et al., 2016; Mollick, 2014; Berns et al., 2020) se sont avérés être déterminants dans les décisions d'investissement. Grâce à toute la littérature qui a été développée jusqu'à présent, autour de ces questions, nous comprenons mieux les décisions des investisseurs qui prennent part au financement participatif. Pourtant, il manque à ce registre de recherches des études qui tiennent compte des sous-groupes qui existent au sein des prêteurs (Lin et Boh, 2019).

Pour contribuer à combler ces lacunes dans la littérature, nous étudions dans ce mémoire les décisions de financement d'un sous-groupe des prêteurs Kiva, les « multi-prêteurs ». En effet, nous avons constaté que ces prêteurs ont contribué au financement de plus de la moitié des projets postés sur Kiva (61,67%) et ont investi en moyenne et au minimum 289 025 $ US (25 $ US x 11 561). En outre, ils viennent pour la plupart (58,06%) des États-Unis, dont la culture valorise plus les motivations intrinsèques qu'extrinsèques (Moonhee et Gawon, 2017). Selon Lin et Boh (2019), les multi-prêteurs sont des prêteurs « actifs » qui jouent un rôle très important car ils sont généralement des prêteurs précoces dont le

soutien favorise celui d'autres prêteurs. Nous choisissons donc d'examiner les stratégies de diversification qu'adoptent ces prêteurs, leur réaction aux facteurs intrinsèques (langage décrivant un prêt comme une opportunité d'aider les personnes dans le besoin) et extrinsèques (langage décrivant un prêt comme une opportunité rentable) contenus dans les récits de demande de prêts fournis par les emprunteurs et enfin leur réaction aux prêts de groupe.

Plus précisément, nous mesurons d'abord le niveau de diversification de chaque portefeuille de prêts et évaluons leur lien avec les choix de financement des prêteurs. Ensuite, nous étudions au regard de l'étude de Allison et al. (2015), l'importance de l'orientation des récits de demande de prêts dans les décisions de financement de ces prêteurs. Enfin, en nous appuyant sur l'étude de Pham et Shen (2016), nous estimons l'effet des prêts de groupe pour ce sous-groupe de prêteurs.

Cette étude devrait apporter un éclairage sur les choix de financement de ces prêteurs qui semblent être très actifs sur la plateforme et ainsi permettre aux emprunteurs de mieux réussir leur levée de fonds.

Le reste de ce travail se présente comme suit. Le premier chapitre donne un aperçu sommaire du crowdfunding, en particulier le prêt pair-à-pair. Nous y abordons les débuts du financement participatif, les formes et les plateformes qui existent. Nous faisons surtout un focus sur le fonctionnement de la plateforme Kiva qui est notre terrain d'étude. Nous détaillons également dans ce chapitre les causes de l'expansion du marché des prêts pair-à-pair ainsi que ses défis. Le deuxième chapitre porte sur la revue de littérature. Nous y présentons une revue de littérature sur la diversification des portefeuilles de prêts, une sur les motivations intrinsèques et extrinsèques, une sur l'importance des récits de demande de prêts et une sur les prêts de groupe. Dans le troisième chapitre, nous formulons les objectifs et les hypothèses de cette étude. Nous abordons la méthodologie dans le quatrième chapitre. La présentation des données et des variables est faite au chapitre 5. Dans les chapitres 6 et 7, respectivement, nous présentons une analyse descriptive des données et les résultats obtenus assortis d'une discussion. Enfin, dans la conclusion de ce mémoire, nous résumons les résultats obtenus, abordons les limites de cette étude et suggérons des pistes pour des études futures.

1. Bref aperçu du marché de prêts pair-à-pair

Cette recherche étant axé sur les prêts pair-à-pair, il importe de clarifier ce concept qui est encore assez récent. Nous faisons donc dans ce chapitre une présentation sommaire de ce marché en partant de ses origines jusqu'aux risques et défis auxquels il est confronté, en passant par les causes de son expansion fulgurante. Nous y détaillons également comment la plateforme Kiva fonctionne.

1.1. Du crowdsourcing au crowdfunding

Le crowdfunding ou financement par la foule est une forme de financement assez récente et en plein essor qui tire sa source du crowdsourcing. En juin 2006, dans Wired Magazine, le terme crowdsourcing a été mentionné pour la première fois par Jeff Howe et Mark Robinson pour désigner l'externalisation par une entreprise d'une tâche ou d'une fonction à priori réservée à ses employés vers un plus grand nombre de personnes (Howe, 2006)[3]. Les entreprises ont donc recours au crowdsourcing pour tirer des idées et profiter de l'expertise de la foule.

Avec le crowdfunding, il s'agit plutôt de tirer profit de la capacité de financement de la foule. Il consiste à réunir une somme d'argent pour financer un projet spécifique grâce à de nombreuses petites contributions de particuliers. Schwienbacher et Laralde (2010) définissent ce moyen de financement comme le financement d'un projet ou d'une entreprise par un groupe d'individus en lieu et place des professionnelles. A la différence donc du financement classique connu, le crowdfunding ne nécessite pas la présence d'institutions financières comme intermédiaires financiers. Pour Belleflamme et al. (2010), le crowdfunding permet de mettre directement les agents à besoins de financement en contact avec les agents à capacité de financement et ce au moyen du web 2.0 (Wenzlaff et al., 2012). Par conséquent, il vise à répondre à la fois aux besoins de financement des porteurs de projets et aux besoins d'investissements des contributeurs tout en éliminant la présence des intermédiaires. Pour répondre à ces objectifs, diverses formes de crowdfunding se sont développées au fil des années. Le financement participatif par la foule est donc généralement subdivisé en quatre catégories : le financement participatif basé sur le don, le financement participatif basé sur la récompense, le financement participatif par action et le

[3] Howe, J. (2006). The Rise of Crowdsourcing. *Wired magazine*

financement participatif basé sur la dette (Schwienbacher et Larralde, 2010; Bradford, 2012 et Lambert et Schwiembacher, 2010). La première forme permet de collecter des fonds pour soutenir des causes caritatives. Selon Glaeser et Shleifer (2001), les projets structurés sous cette forme ont tendance à réussir plus facilement leur collecte de fonds car ils sont plus susceptibles d'offrir des produits de haute qualité contrairement aux organisations à but lucratif qui sont soumises à la maximisation des bénéfices. Le financement participatif basé sur la récompense, quant à lui, permet aux contributeurs d'obtenir une récompense qui est le plus souvent non financière en échange de leur investissement. On classe également dans cette catégorie le financement participatif par préachat pour lequel la récompense est un produit, gratuit ou à prix réduit, de l'entreprise financée. Le financement participation par action, en impliquant la vente de titres, permet aux investisseurs de détenir une partie du capital d'une entreprise en contrepartie de leur investissement. Ils peuvent donc recevoir une partie du bénéfice ou du rendement des projets dans lesquels ils investissent ainsi qu'un droit de vote. Cette forme de financement est beaucoup plus complexe à cause de la nécessité de valoriser l'entreprise et la non-implication de l'investisseur dans ce processus de valorisation. L'investisseur investit donc dans un titre financier sans la possibilité de participer à sa négociation.

Dans le cadre de cette étude, nous nous intéressons spécialement au financement participatif basé sur la dette généralement appelé peer-to-peer lending, en français le prêt pair-à-pair. Le prêt P2P est le type le plus populaire de financement participatif et le terme pair-à-pair qui compose son nom décrit une interaction directe entre particuliers. Il peut donc être défini comme l'utilisation d'une plateforme en ligne qui met directement en relation de potentiels prêteurs avec des emprunteurs ayant besoin d'un accès au capital. Pour certains de ces prêts, le principal doit être remboursé avec des intérêts tandis que pour d'autres, seul le principal est remboursé aux prêteurs.

1.2. Le financement participatif basé sur la dette

L'industrie du prêt pair-à-pair a démarré en février 2005 avec l'apparition de Zopa.com au Royaume-Uni mais tire sa source des principes de la microfinance dont les débuts remontent à la création de la Grameen Bank dans les années 70. Dans les années qui ont suivi le lancement de Zopa, de nombreuses autres plateformes ont été rapidement mises en place au point où l'on comptait dix ans plus tard pas moins de 820 plateformes P2P (CCAF-FCVI, 2016a à 2016d). En termes de prêts accordés, ce marché n'a pas cessé de s'agrandir. En 2015, le volume mondial des prêts accordés sur ce marché s'évaluait à 25,1 milliards de dollars, ce qui représente 72,88% de celui du marché de financement participatif dans son ensemble[4]. Selon Moldow (2015)[5], ce marché des prêts pair-à-pair atteindrait 1 000 milliards de dollars de prêts d'ici 2025. Les prêts consistent généralement en de petites contributions provenant d'un grand nombre de prêteurs et sont souvent à court ou moyen terme (trois à cinq ans) pour des objectifs tels que la consolidation de cartes de crédit ou de dettes, les prêts automobiles ou l'acquisition d'un logement ou encore l'entrepreneuriat. Contrairement aux prêts conventionnels, ces prêts sont généralement non garantis pour la simple raison que les plateformes de prêts pair-à-pair servent simplement de médiateur entre les emprunteurs et les prêteurs, et qu'aucune banque n'intervient dans ce système. Par conséquent, ces prêts ne sont pas adossés à une garantie ni couverts par une assurance-dépôts.

Bien que son origine soit la microfinance dont l'objectif principal est d'aider les personnes démunies, ce secteur est dominé par des organisations à but lucratif, telles que Prosper et Lending Club, qui facturent des intérêts sur les prêts pair-à-pair. Pour un prêt donné, le taux d'intérêt appliqué par ces plateformes est généralement supérieur aux taux d'épargne disponibles pour le prêteur, mais inférieur au taux d'intérêt appliqué pour les prêts non garantis sur le marché traditionnel. Certaines de ces plateformes utilisent un modèle d'enchères pour déterminer le taux d'intérêt. Dans un tel cas, ce sont les prêteurs eux-mêmes qui déterminent le taux d'intérêt d'un prêt spécifique en se basant sur les informations fournies par l'emprunteur sur sa cote de crédit, ses revenus, ses arriérés antérieurs, l'objet du prêt et des informations plus personnelles comme leur photo, leur

[4] http://crowdexpert.com/crowdfunding-industry-statistics/
[5] P2P-LENDING-PRINT-H-v2-r10.indd (foundationcapital.com)

origine ethnique, leur lieu de résidence, etc. Cependant, la plupart des principales plateformes fixent les taux d'intérêt en fonction de leur propre évaluation du risque de l'emprunteur.

Fondée en 2005 à San Francisco, Kiva est la principale plateforme à but non lucratif qui permet à des emprunteurs de divers pays de bénéficier de prêts sans intérêt. Visant à réduire la pauvreté, les prêts accordés sur Kiva sont principalement destinés aux entrepreneurs des pays en développement. A date, Kiva a facilité la collecte et le transfert de 1,61 milliard de dollars de prêts sans intérêt, financés par ses 1,9 millions de prêteurs pour 4 millions d'emprunteurs dans 77 pays sur 5 continents.

Au nombre de 3 140[6], les partenaires de terrain sont des institutions de microfinance locales dont le rôle consiste à sélectionner et évaluer les emprunteurs avant que leurs demandes de prêts ne soient publiées sur le site de Kiva. Ils travaillent directement avec les emprunteurs pour fixer les montants et les conditions des prêts, décider de la nécessité de constituer les emprunteurs en groupe, recueillir leurs informations (par exemple des informations sur l'objectif du prêt, le montant recherché, le partenaire de terrain le cas échéant, l'identité de l'emprunteur et son histoire d'entreprise etc.) et les télécharger sur le site de Kiva.

Kiva a, de son côté, la charge de sélectionner, noter et surveiller chaque institution de microfinance sur sa plateforme et lui attribuer une note de risque que les prêteurs doivent prendre en compte dans leurs décisions de financement. Une fois la demande publiée et financée, Kiva rassemble les fonds des prêteurs et les transmet aux IMF, qui accordent les prêts aux emprunteurs et gèrent les remboursements. Les partenaires de Kiva facturent aux emprunteurs des intérêts sur les prêts afin d'assurer leurs coûts de fonctionnement mais les responsables de Kiva affirment qu'ils veillent à ce que ces taux d'intérêt ne soient pas exorbitants. Par conséquent, lorsque les partenaires de terrain collectent les remboursements, ils retirent les intérêts et autres frais avant de transférer le principal du prêt à Kiva. La transmission des remboursements est assurée par Kiva qui crédite le compte de chaque prêteur. Comme toute plateforme P2P, Kiva n'assure pas les fonctions d'intermédiaire telles que la transformation du risque ou des échéances. Ainsi, lorsqu'un partenaire informe Kiva qu'un emprunteur n'a pas effectué un paiement, Kiva informe à son tour les prêteurs qui recevront soit un paiement tardif ou partiel ou pire encore ne

[6] https://www.Kiva.org/about/where-Kiva-works

recevront aucun paiement. En dépit de cette réalité de risque, Kiva connaît un taux de remboursement d'environ 96,3%.

Même si la majorité des demandes de prêts effectuées sur Kiva transite par leurs partenaires de terrain, certains emprunteurs effectuent leur demande directement sur le site sans l'aide d'aucun partenaire de terrain. Dans un tel cas, l'emprunteur est contraint de collecter une partie du prêt auprès de ses amis et de sa famille ou solliciter l'appui d'un syndic approuvé par Kiva pour lui servir de garant et n'est financé qu'à condition que le montant collecté à l'issu de sa campagne soit au moins égal au montant recherché. Quant aux demandes indirectes, c'est-à-dire celles qui transitent par des partenaires, elles peuvent être pré-décaissées ou post-décaissées. La plupart de ces prêts sont pré-décaissés, c'est-à-dire versés aux emprunteurs avant même que les détails de ladite demande de prêt ne soient publiés sur le site Web de Kiva. Dans le cas des prêts post-décaissés, les emprunteurs ne reçoivent le prêt qu'après son financement total sur la plateforme. Le financement collecté sur Kiva sert donc généralement à réapprovisionner les partenaires pour les prêts qu'ils ont, en fait, déjà déboursés pour répondre aux besoins des emprunteurs.

Kiva applique deux modèles de financement : le flexible et le fixe. Le modèle fixe n'est rien d'autre que le tout-ou-rien dans lequel le montant cible doit être atteint avant que les fonds ne soient envoyés au partenaire ou à l'emprunteur. Tous les prêts directs suivent ce modèle ainsi que les prêts qui ne sont pas pré-décaissés par les partenaires de terrain.

Dans le modèle flexible, comme c'est le cas dans le keep-it-all, les emprunteurs peuvent garder les fonds collectés même si le montant ciblé n'est pas atteint. Dans ce cas, le partenaire est chargé de trouver d'autres sources de financement afin de compléter le reste du prêt.

Les prêteurs ont une page de profil sur le site de Kiva et peuvent fournir des renseignements telles que leur nom et leur adresse électronique ou décider de rester anonyme. Ils ont accès à la liste des demandes de prêts, qu'ils peuvent consulter pour sélectionner les entrepreneurs qu'ils souhaitent financer. Contrairement aux prêteurs des plateformes P2P à but lucratif, les prêteurs de Kiva n'accordent des prêts que pour le plaisir d'aider et faire une différence dans la vie d'une personne. Pour ce faire, ils ont la possibilité de débourser seulement 25 dollars ou plus, et pourraient encore utiliser les remboursements pour aider d'autres. Ils disposent, en règle générale, de 30 jours pour financer les projets de

leur choix sur la base des informations sur le prêt et l'emprunteur et durant toute la période que durera le prêt, ils reçoivent périodiquement des mises à jour par courriel sur le développement de l'entreprise dans laquelle ils ont investi.

1.3. Les motivations à la participation au marché de prêts pair-à-pair

Étant donné la croissance impressionnante que connaît le marché du P2P, il importe d'aborder les raisons de cette évolution car elles renseignent également sur les motivations des utilisateurs. Une raison fondamentale et évidente est le développement d'internet. Selon Kirby et Worner (2014), le financement participatif a commencé son ascension seulement après l'innovation technologique des applications Web 2.0 sur Internet. Cette avancée technologique a permis la mise en place de plateformes de financement participatif pratiques et simples. En plus de la commodité, les utilisateurs de ces plateformes en ligne apprécient la qualité et la rapidité des services offerts. Ce progrès a aussi eu pour conséquence la réduction des coûts d'exploitation et l'ouverture de ce marché à un public plus vaste (Klafft, 2008a et 2008b).

Une autre raison que les auteurs soulèvent est la crise de 2008. La crise financière mondiale a laissé place à un environnement dans lequel les banques et les prêteurs traditionnels ont vu leur capacité de prêts être limitée par les nouvelles réglementations. Par exemple, le montant annuel de prêts bancaires aux États-Unis qui était de l'ordre de 1,75 trillion en 2007 est passé à 0,65 trillion en 2008 puis à 0,45 trillion en 2009. Ce contexte a révélé le marché P2P comme une alternative aux institutions de prêt traditionnelles pour les particuliers et petites entreprises qui ne satisfaisaient pas aux critères plus stricts d'octroi de prêts traditionnels et même pour certains particuliers qui présentaient un risque raisonnable (Packer, 2010). Aujourd'hui encore, le marché P2P gagne en popularité auprès des entreprises en démarrage et des particuliers qui sont exclus des prêts traditionnels. Pour ces emprunteurs, le marché de prêts P2P assure un meilleur accès au crédit.

D'autres emprunteurs, notamment ceux à faibles revenus, sont attirés par les prêts P2P à cause du coût du capital faible par rapport aux banques (Wang et al. 2009). Cette différence de coût en faveur du prêt P2P résulte de la nature ciblée des activités des plateformes en ligne et de la réduction de la bureaucratie ainsi que de la documentation requise sur ces plateformes par rapport aux banques. Aussi, contrairement aux banques, les plateformes ne sont pas tenues de détenir des capitaux réglementaires.

Selon Estrin et al. (2018), participer à un tel marché permet aux entrepreneurs d'accéder rapidement à des niveaux élevés de capitaux et de faire connaître leur entreprise ou produit à un public plus large. En effet, les start-ups peuvent utiliser les plateformes de P2P pour se faire de la publicité en bénéficiant de stratégies de markéting en occurrence celle du « bouche à oreille ». Ils peuvent également bénéficier du réseau des investisseurs ainsi que de leurs compétences et connaissances pour leurs projets (Cumming et Johan, 2013; Ordanini et al., 2011). Certains entrepreneurs peuvent aussi, à la suite d'un financement par la foule, bénéficier des fonds du marché financier traditionnel (Cumming et Johan, 2013). La plupart des emprunteurs apprécient la flexibilité que leur offrent les plateformes de prêts P2P de pouvoir annuler à tout moment les contrats de prêt sans payer de pénalité de remboursement anticipé.

Pour les prêteurs, leur participation à ce marché est liée à la possibilité d'obtenir des taux de rendements beaucoup plus élevés que ceux disponibles pour les prêts conventionnels. Avec des taux de dépôts réels parfois négatifs, nombreux sont les épargnants qui décident d'investir dans le prêt P2P. Selon le Wall Street Journal, à une époque où les rendements étaient au plus bas, les investisseurs en prêts P2P ont bénéficié de rendements annuels allant jusqu'à 10 % ou plus. Les prêts P2P représentent également pour les prêteurs un nouveau produit leur permettant de diversifier davantage leur portefeuille. Sur les plateformes de prêts P2P, les prêteurs disposent de toutes sortes de prêts et peuvent sélectionner des prêts qui correspondent à leur niveau d'aversion pour le risque et le rendement souhaité. Ils peuvent diversifier leur portefeuille en investissant des montants relativement petits dans des prêts à la consommation et/ou commerciaux, ce qui n'est pas possible avec les instruments d'épargne traditionnels. En outre, le crowdlending offre aux prêteurs pour qui l'aspect social est important, la possibilité d'interagir avec les emprunteurs et d'allouer eux même leur argent.

Au-travers des crédits accordés aux entrepreneurs, en particulier aux PME, le marché P2P est en train de devenir un moteur important pour la croissance économique dans les pays concernés. C'est ce qui explique le soutien de certains gouvernements à l'égard de ce marché comme le gouvernement britannique, qui après la crise financière a utilisé ces plateformes pour attribuer des prêts aux PME financés par l'État. Cependant, le marché du

prêt P2P ne présente pas que des avantages. Les utilisateurs de ces plateformes en ligne font face à des risques qui justifient la réticence de certains d'y participer (Estrin et al., 2018).

1.4. Les risques et défis du marché des prêts pair-à-pair

D'un point de vue juridique, le prêt pair-à-pair se résume en un contrat de prêt (dette) qui implique la promesse de rembourser le montant principal et les intérêts, le cas échéant, dans le futur. L'incertitude liée à cette promesse rend donc les contrats de prêt très complexes et inévitablement incomplets, puisque toutes les circonstances possibles ne peuvent être couvertes (Davis, 1995). Par conséquent, les prêts P2P héritent de cette nature risquée des contrats de prêts mais en plus présentent des risques inhérents à l'environnement en ligne ainsi que des facteurs de risque liés aux caractéristiques des emprunteurs qui sont pour la plupart non bancales. Les risques que peuvent représenter les prêts P2P pour les emprunteurs, les prêteurs et les opérateurs de plateforme ont été analysés dans la littérature. On y distingue entre autres: le risque de crédit, le risque opérationnel, le risque de liquidité, l'asymétrie d'information, le risque de fraude et de cyber-attaque et le dommage réputationnel lié à l'échec d'un financement.

Le risque de crédit ou risque de contrepartie ou encore risque de défaillance est propre à tout prêt ou investissement. Il résulte de l'incapacité d'un emprunteur à honorer le contrat de prêt parce qu'il est négligent ou que ses moyens sont insuffisants pour rembourser le prêt. Cela révèle l'importance d'une évaluation correcte de la capacité des emprunteurs à rembourser les prêts. Bien que les plateformes de prêts P2P utilisent les informations des rapports de crédit des emprunteurs dans l'évaluation de leur solvabilité, elles ne vérifient souvent pas les informations fournies par les emprunteurs, telles que leur revenu, leur ratio dette/revenu, leur emploi et leur statut de propriétaire. De plus, les prêts P2P sont ficelés sur la base de garanties sociales ce qui fait qu'ils attirent les emprunteurs ayant une mauvaise cote de crédit (Bruett 2007). A ces débuts en 2006, la plateforme Prosper a enregistré un taux de défaillance très inquiétant de 30%. Cette situation a conduit à la mise en place de critères d'investissement plus stricts, grâce auxquels le taux de défaillance global de ce secteur en 2013 se situerait entre 0,2 % et 7 % (Kirby et Worner, 2014). Toutefois, le risque de crédit sur ce marché demeure non négligeable.

Le risque opérationnel peut toucher à la fois les prêteurs et emprunteurs sur le marché de prêts P2P. Ce risque se produit en cas de pertes financières inattendues dues à

l'inadéquation ou à l'échec des processus internes de gouvernance des plateformes. Il peut s'agir, par exemple, du processus de sélection des demandeurs de prêts ou celui du recouvrement des prêts. Dans le cas du recouvrement des prêts, les prêteurs dans les prêts P2P n'ont pas de recours direct contre les emprunteurs et doivent compter sur les plateformes et leurs agences de recouvrement (Chaffee et Rapp, 2012). Pour ce qui est de la sélection des demandeurs, les emprunteurs peuvent faire l'objet d'une évaluation incorrecte de la part de la plateforme ce qui peut réduire leur chance de susciter l'attention des prêteurs potentiels.

Le rendement attendu élevé que propose le marché de prêts P2P peut amener certains investisseurs à négliger le risque d'illiquidité important qui existe avec ces prêts. Il est question d'un risque qui survient lorsque le prêteur n'a pas de possibilité de vendre le contrat, sauf à une forte décote par rapport à la valeur nominale, une fois l'argent investi. Étant donné qu'il n'existe pas de marché secondaire pour les billets de prêts P2P, ce risque est non négligeable pour ce type d'investissement (Kirby et Worner, 2014). De plus lesdits prêts sont tout au plus à moyen terme (Chaffee et Rapp, 2012).

Le risque le plus important sur ce marché est celui de la sélection adverse et de l'aléa moral car il existe toujours une asymétrie d'information entre les emprunteurs et les prêteurs (Freedman et Jin, 2011). Sur les plateformes où les taux d'intérêts sont appliqués, ces taux sont les seules informations concrètes dont disposent les prêteurs sur les emprunteurs. Pour différencier un mauvais emprunteur d'un bon, sur les plateformes comme Kiva où des intérêts ne sont pas appliqués, les prêteurs ne disposent que d'informations non vérifiées sur les emprunteurs et doivent donc faire preuve d'une diligence raisonnable et placer une confiance aveugle dans la capacité des plateformes et des partenaires de terrain à évaluer convenablement le risque (Kirby et Worner, 2014).

A ces risques, il faut ajouter le risque de fraude et de cyber-attaque. A cause de l'anonymat que requiert Internet, les plateformes de financement participatif présentent des risques plus élevés de fraude allant de l'usurpation d'identité au blanchiment d'argent. La plupart des plateformes s'assurent d'effectuer un certain nombre de vérifications en ce qui concerne les identités des emprunteurs, mais pas pour les prêteurs (Kirby et Worner, 2014).

Contrairement à la perception générale, les niveaux potentiels de capital disponibles sur les plateformes de financement participatif sont plutôt limités, et même si ce type de

financement offre aux entreprises la possibilité de recueillir des fonds auprès d'une large plage d'investisseurs, de nombreuses campagnes ont tendance à rester infructueuses (Lukkarinen et al., 2016). Estrin et al. (2018) relèvent que l'échec d'une campagne de financement sur ces plateformes a un important dommage réputationnel car l'échec est public, ce qui non seulement ferme la porte au financement traditionnel mais constitue un signal public négatif pour les futurs investisseurs.

2. Revue de littérature

Dans ce chapitre nous survolons les études qui ont été menées avant celle-ci sur la question de la diversification des portefeuilles de prêts en crowfunding (section 2.1), les motivations des crowfunders (section 2.2), l'importance des récits de demande de prêts (section 2.3) et les groupes de prêts (section 2.4). Cet exercice a pour but de montrer la nécessité du travail qui sera réalisé dans ce mémoire (section 2.5).

2.1. Revue de littérature sur la diversification des portefeuilles de prêts

L'un des importants avantages du financement participatif réside en ce qu'il offre une grande variété d'investissements, permettant ainsi aux investisseurs d'avoir un grand niveau de diversification. Selon la théorie financière classique, la diversification est une règle de base nécessaire à l'optimisation de tout portefeuille d'investissement. En effet, les travaux menés par Merton (1969) et Samuelson (1969) sur le comportement des investisseurs ont conduit à un résultat capital qui prouve qu'en général, et contrairement au rendement, le risque d'un portefeuille n'est pas égal à la moyenne des risques des actifs inclus dans ce portefeuille. Cela suggère que plus un investisseur effectue de nombreux investissements et plus le risque global de son portefeuille par rapport au risque moyen des actifs qui composent le dit portefeuille diminue. Par conséquent, quel que soit son niveau d'aversion au risque, tout investisseur rationnel est censé détenir un portefeuille diversifié. La plupart des travaux dans le domaine de la diversification portent sur le marché des actions et démontrent que la diversification est un moyen efficace pour contrôler l'exposition au risque non systématique (Haugen, 2001; Diamond, 1984; Knill, 2009 ; Norton et Tenenbaum, 1993; Goetzmann et Kumar, 2008). Toutefois, d'autres études portant sur le marché bancaire aboutissent à des résultats qui remettent en cause cette efficacité (DeYoung et Roland, 2001; Stiroh, 2004; Stiroh et Rumble, 2006). Ces études attribuent ce résultat au fait que les revenus autres que les intérêts générés par les banques fluctuent beaucoup et tendent à annuler les avantages de la diversification.

Dans le domaine du financement participatif, il existe très peu d'études qui traitent de la question de la diversification des portefeuilles de prêts. Au-delà du fait que le financement participatif soit un phénomène récent, la difficulté qu'il y a à identifier les investisseurs est

l'une des raisons de ce vide dans la littérature (Harrison et Mason, 1992; Landström, 1993; Wetzel, 1983; Erikson et al., 2003). En effet, les investisseurs ont tendance à préférer l'anonymat ou ne fournissent que très peu d'informations sur leur profil. Pour surmonter cet obstacle, Mintjes (2016) a utilisé à la fois les données sur les ménages dont dispose la Banque centrale néerlandaise et les données de prêts de 5 plateformes de financement participatif, ce qui lui a permis d'obtenir des informations fiables sur 863 investisseurs. Joo-Kitano (2021), quant à lui, a dû cibler les investisseurs de plusieurs pays (Allemagne, Pays-Bas et Espagne) avant de parvenir à interroger 242 investisseurs. Dans ces deux études, l'objectif des auteurs a été, entre autres, de rechercher les déterminants de la diversification des portefeuilles en crowdfunding. En particulier, Mintjes (2016) a trouvé qu'une part importante (27,82% de l'échantillon) des ménages néerlandais détient des portefeuilles de crowdfunding fortement sous-diversifiés et que la littératie financière affecte positivement le niveau de diversification des portefeuilles. Joo-Kitano (2021), de son côté, trouve que l'éducation formelle, le statut professionnel et le profil de l'investisseur évalué expliquent significativement la stratégie de diversification des risques dans le financement participatif par action. Dans son échantillon seulement 4% des investisseurs détient des actions de plus de 5 compagnies tandis que 57% n'en détient qu'une seule. Ces quelques résultats révèlent une sous diversification des portefeuilles de prêts dans le financement participatif malgré les possibilités infinies d'investissements qu'offre ce marché.

2.2. Revue de littérature sur les motivations intrinsèque et extrinsèque

Les recherches sur les motivations des individus ont conduit à la distinction de deux grandes catégoriques de motivations : les motivations intrinsèques et les motivations extrinsèques (Vallerand, 1997). Les motivations d'un individu à réaliser une activité ou une tâche sont dites intrinsèques lorsqu'il réalise l'activité ou la tâche en question pour le simple plaisir inhérent ou la satisfaction qu'elle peut lui procurer tandis qu'elles sont dites extrinsèques lorsqu'il la réalise dans le but d'obtenir un résultat positif ou éviter un résultat négatif qui est dissociable de l'activité, par exemple obtenir une récompense ou éviter une punition (Deci, 1975; Venkatesh, 1999). Dans le contexte des prêts en ligne, ces deux types

de motivations coexistent et vraisemblablement sous-tendent le choix d'investissement de la foule.

Concernant les plateformes à but lucratif, l'acception générale est que la décision d'investissement est positivement corrélée aux avantages monétaires (récompenses extrinsèques). En étudiant les motivations à investir sur les plateformes de financement participatif basé sur la récompense et les actions, Cholakova et Clarysse (2015) ont observé que la décision d'investissement est positivement associée au besoin de récompense des investisseurs et à leur désir de faire confiance aux emprunteurs. Les auteurs démontrent que le désir d'aider les autres, de soutenir une idée ou d'appartenir n'influence pas les décisions d'investissement. Spécifiquement, les individus qui décident d'investir dans le financement participatif basé sur les actions sont motivés par les rendements financiers (récompenses extrinsèques) qu'ils peuvent en tirer. Ces investisseurs semblent donc avoir des motivations extrinsèques à investir dans le prêt social basé sur les actions (Collins et Pierrakis, 2012). Néanmoins, Bretschneider et Leimeister (2017) trouvent que des motivations non financières, telles que la reconnaissance des autres et le fait de contribuer à la réalisation d'un projet particulier, sont aussi positivement associées à la décision de participer au financement participatif basé sur les actions.

Sur les plateformes de financement participatif à but non lucratif, telles que Kiva, les motivations répertoriées sont tout du moins assez différentes. McKinnon et al. (2013), dans leur analyse sur les motivations des prêteurs Kiva, dénotent une similitude particulière avec les discours du néolibéralisme. Ils identifient trois principales catégories de motivations : (a) la promotion du travail acharné (« *I want to provide economic opportunity for hard-working people around the world* »), (b) l'incitation à l'entrepreneuriat (« *the microlending method empowers the entrepreneur to grow and develop their business enabling them to succeed on their own*»), et (c) la responsabilité (« *I would like to help to make a difference to those who are willing to take a chance to better themselves* »). Liu et al. (2012), quant à eux, en identifient plutôt dix dont l'altruisme général, l'altruisme spécifique à un groupe, l'empathie, la réciprocité, la satisfaction personnelle, le devoir religieux et les raisons externes. Mittelman (2013) a rangé en fonction de leur popularité, les motifs fournis par les prêteurs. Nous trouvons premièrement le désir de faire une différence dans la vie des familles; ensuite l'altruisme, la justice sociale et la culpabilité existentielle et enfin

l'égoïsme, la sympathie, l'engagement religieux et l'empathie. La décision de ces investisseurs de contribuer aux prêts pair-à-pair semble être essentiellement basée sur des motivations intrinsèques.

2.3. Revue de littérature sur les récits de demande de prêt

La question de l'importance des informations fournies dans les récits de demande de prêt et de leur lien avec la décision des prêteurs de contribuer au financement d'un projet plutôt qu'un autre a été abordée dans plusieurs travaux empiriques et sur différentes plateformes.

Par exemple, à partir de 220 257 demandes de prêts publiées sur la plateforme Prosper, Larrimore et al. (2011) ont montré que la longueur d'un récit de prêt impacte l'issu du processus de financement. Dans les commerces en ligne (par exemple eBay), la longueur de la description des produits est associée à une augmentation des enchères et des prix de ventes (Flanagin, 2007). Les résultats prouvent donc que, tout comme dans ces types de commerces, les récits de demande de prêts plus longs contribuent à réduire l'incertitude sur le marché de prêts pair-à-pair, ce qui se manifeste par une augmentation de la probabilité de succès de la collecte de fonds. Toutefois, l'ajout de mots pour compléter un court récit est plus bénéfique que l'ajout de mots à un récit déjà long. Ils montrent également que l'utilisation de quantificateurs (par exemple plusieurs, beaucoup, peu, etc) et de prépositions (par exemple dans, à, de, sur, avec, etc) est positivement associé au succès des campagnes de financement car ils rendent les récits plus concrets. Les informations quantitatives présentes dans les récits de demande de prêts sont considérées par les prêteurs comme un indicateur de la capacité de remboursement des emprunteurs et donc augmentent la probabilité de réussite des collectes de fonds. Les auteurs ont mesuré ces informations quantitatives à partir de mots numériques (par exemple deuxième, deux, mille, etc) et de mots d'argent (par exemple argent comptant, devoir, audit, argent, devoir, etc). En revanche, les détails humanisants (par exemple les informations sur la famille, les réalisations, les activités de loisir, la maison ou la religion) et les justifications que donnent les emprunteurs dans les récits sont négativement associés au succès des campagnes de financement. Ces informations semblent être perçues par les prêteurs comme des informations non pertinentes pour l'évaluation de la capacité des emprunteurs à rembourser les prêts.

Dans les récits de demande de prêts, certains emprunteurs revendiquent des identités qui impactent significativement la décision des prêteurs. Herzenstein et al. (2011) ont identifié 6 types de revendications d'identité: emprunteur digne de confiance, emprunteur ayant une entreprise prospère, emprunteur honnête, emprunteur travailleur, emprunteur rencontrant des difficultés économiques et emprunteur religieux. Les auteurs ont montré que le fait qu'un porteur de projet se décrive comme digne de confiance ou possédant une entreprise prospère impacte positivement la décision de financement des prêteurs tout comme le fait de fournir des détails relatifs aux activités de travail (Larrimore et al., 2011). Cependant le fait de se présenter comme ayant des difficultés économiques ou de fournir des détails sur sa famille, ses activités sportives, sur la nécessité du prêt ou encore sur sa religion l'impacte négativement.

En outre, Parhankangas et Renko (2017) ont montré que l'importance du style linguistique varie en fonction de l'emprunteur en comparant les campagnes des entrepreneurs sociaux à celles des entrepreneurs commerciaux. Ces derniers sont des porteurs de projet qui cherchent à développer de nouveaux produits ou services pour les marchés de consommation alors que les entrepreneurs sociaux ont pour objectif d'apporter un bien social, plutôt que d'accroître leur richesse personnelle ou celle des actionnaires (Dacin et al., 2010 ; Renko, 2013). Les auteurs ont défini quatre styles linguistiques à partir des données : le langage précis, le langage concret, le langage interactif et le langage de distanciation psychologique. Le langage concret a été opérationnalisé comme la somme du nombre d'articles, de prépositions et de quantificateurs (Larrimore et al. (2011). Le langage précis a été mesuré en prenant l'opposé du ratio type-token de Johnson, soit l'opposé du rapport entre le nombre de mots différents dans un passage et le nombre total de mots du passage. L'utilisation de questions dans les récits fait montre d'une interaction avec la foule (Tausczik et Pennebaker, 2010). La distanciation psychologique est captée par les mots d'émotion (par exemple blessé, inquiet, triste, chagrin). Les résultats montrent que contrairement aux entrepreneurs commerciaux pour qui seul le langage précis et celui interactif sont associés à une augmentation de la probabilité de réussite, les entrepreneurs sociaux ont intérêts à adopter les quatre styles linguistiques dans leurs récits de demande de prêts. En effet, les entrepreneurs sociaux se doivent d'expliciter le but de leurs projets de façon compréhensive dans leurs récits et de démontrer qu'ils sont proches de la foule car

cela est associé à une augmentation significative de la probabilité de réussite de leur levée de fonds.

L'étude menée par Allison et al. (2015) prouve que l'orientation des récits de demande de prêts déterminent l'issu des levées de fonds dans le financement participatif. Leurs résultats démontrent que les récits avec un niveau élevé de langage extrinsèque sont associés à une durée de campagne plus longue tandis que les récits avec un niveau élevé de langage intrinsèque sont associés avec une durée plus courte. Ils mesurent le langage extrinsèque à partir des mots de profit et de risque qui ont tendance à donner à un projet un aspect commercial, et le langage intrinsèque à partir de mots de diversité et d'intérêt humain qui sont censés donner à un projet un aspect d'aide. Ces conclusions rejoignent celles de Moss et al. (2017) qui ont classé les demandes de prêts publiées sur Kiva en 2012 sous deux catégories linguistiques afin d'examiner comment elles sont associées à la vitesse de financement des projets. En fait, en suivant le processus à plusieurs étapes suggéré par Short et al., (2009, 2010), et Payne et al. (2011), les auteurs ont développé une liste de 179 mots avec 108 mots mesurant l'orientation sociale et 71 l'orientation économique dans les récits. Ils considèrent un projet comme appartenant à une catégorie linguistique où le langage économique est mis en avant lorsque son récit de prêt contient par exemple les mots revenu, riche, salaire, etc. Pour la catégorie linguistique où le langage social est mis en avant, les récits contiennent à la fois des mots à caractère social (santé, aide, humanité, humain, immunité) et des mots en lien avec l'environnement (air, biofertilisant, biogaz, carbone, climat, conservation). Les résultats révèlent que les récits à orientation sociale sont associés à un financement plus rapide que les récits à orientation économique. Cependant, les projets qui présentent une orientation mixte c'est-à-dire les projets à la fois sociale et économique se financent lentement.

2.4. Revue de littérature sur les prêts de groupe

En microfinance, le suivi après signature d'un contrat de prêt, tel que pratiqué dans le cadre des prêts bancaires, est moins efficace à cause de la taille bien trop petite des micro-prêts et du coût élevé du suivi. Néanmoins, ce suivi existe et est mis en œuvre au travers des prêts à responsabilité conjointe ou prêts de groupe. La performance financière de la Grameen Bank et celle de plusieurs autres programmes de micro prêts est essentiellement due à la connaissance que les emprunteurs d'un même groupe de prêt ont de leurs pairs.

Plusieurs recherches théoriques ont montré que contrairement aux prêts individuels, les prêts de groupe permettent de contrer les problèmes de sélection adverse et d'aléa moral grâce à la surveillance par les pairs (Stiglitz, 1990; Banerjee et al., 1994; Ghatak et Guinnane, 1999; Ghatak, 2000). En accord avec cela, on devrait s'attendre à observer une réduction du temps de financement des prêts de groupe et une probabilité de réussite du financement plus élevée pour ces prêts. Cependant, les preuves de cela dans la littérature sont mitigées (Armendariz de Aghion 1999 ; Paxton et al. 2000).

Avec 289 329 demandes de prêts publiées entre le 27 novembre 2007 et 18 juin 2008, Galak et al. (2011) trouvent que les prêteurs préfèrent prêter aux emprunteurs seuls plutôt qu'à un groupe d'emprunteurs. De façon précise, les résultats montrent que le montant accordé par les prêteurs diminue et que le temps de financement augmente en fonction du nombre d'emprunteurs effectuant la demande. Les auteurs expliquent ce résultat en se basant sur l'effet de la victime identifiable ou victime unique selon lequel un emprunteur unique (victime unique ou identifiable) ayant besoin d'un soutien financier suscite une réponse émotionnelle plus forte qu'un groupe d'emprunteurs (victime non identifiable) (Kogut et Ritov, 2005a et 2005b; Small et Loewenstein, 2003).

À l'opposé, Berns et al. (2020) montrent, en utilisant 146 218 demandes de prêts publiées entre 2008 et 2013 sur Kiva, que la probabilité qu'un projet soit totalement financé augmente lorsqu'il s'agit d'un projet de groupe. Les résultats obtenus par Ly et Mason (2012b), avec les 132 495 observations collectées entre 2006 et 2009, révèlent que la préférence des prêteurs pour les groupes d'emprunteurs par rapport aux emprunteurs individuels dépend du nombre d'emprunteurs qui composent les groupes d'emprunteurs. Les auteurs mettent en lumière le fait que les prêts individuels se financent 84% plus rapidement que les prêts de groupe ayant une moyenne de 1,7 membres. En revanche,

lorsque le nombre de membre est d'au moins 7, les prêts de groupe sont préférés aux prêts individuels. Ces résultats révèlent l'importance qu'accordent les prêteurs à l'impact potentiel de leurs contributions. En fait, il est possible que les prêteurs aient l'impression qu'en finançant un prêt à un grand groupe, leur contribution touche plusieurs personnes qu'en se concentrant sur une seule personne, ce qui conduit à une augmentation de la vitesse de financement à mesure que le nombre de membres du groupe augmente. Cette idée est soutenue par Pham et Shen (2017) qui montrent, à partir d'un échantillon de 995 911 demandes de prêts, qu'en moyenne lorsque les emprunteurs se constituent en groupe, le temps de financement se trouve être réduit de 3,3 jours. Pour eux, combiner au fait que les prêteurs Kiva peuvent prêter le montant qu'ils veulent (25 $ minimum) et peuvent diversifier leur investissement en faisant de petits prêts dans plusieurs projets, ce résultat prouve que les prêteurs Kiva accordent plus d'importance au nombre de vies qu'ils peuvent toucher par leur contribution plutôt qu'à la probabilité de remboursement plus élevée qu'offre les prêts de groupe. La particularité de l'étude de Pham et Shen (2016) par rapport aux autres études réside dans la prise en considération des récits de demande de prêts dans leur analyse. En effet, l'inexpérience des prêteurs en ligne en matière d'évaluation du niveau de solvabilité des emprunteurs potentiels amplifie les problèmes d'asymétrie d'information sur le marché du financement participatif, ce qui fait des récits de demande de prêts une mine d'informations cruciales pour ces prêteurs. Même si Berns et al. (2020) ont aussi utilisé ces récits dans leur étude, ils n'ont pris en compte que les informations relatives à l'altruisme en mesurant la fréquence des mots altruistes que ces récits comportent. L'étude de Pham et Shen (2016) est donc la première à avoir pris en considération, dans l'analyse des prêts de groupe en financement participatif, l'intégralité des informations que pourraient contenir les récits de demande de prêts.

A notre connaissance, l'étude la plus récente qui analyse également les prêts de groupe sur la plateforme de Kiva est celle de Priberny et Sommer (2021). Les auteurs ont utilisé 1 837 525 observations allant du 1er janvier 2009 au 24 juillet 2020 pour montrer que les prêts de groupe ont une probabilité plus élevée d'être totalement financés comparé aux prêts individuels mais prennent plus de temps pour leur financement. Ils trouvent que la probabilité de réussite des campagnes de financement des prêts de groupe est 1,023 fois celle des prêts individuels et que les projets de groupe qui réussissent leur campagne de

financement mettent 1,022 seconde de plus que les projets individuels pour y arriver. Même s'il ne s'agit que de quelques secondes de différence, cet écart est quand même significatif.

2.5. Conclusion de la revue de littérature

En résumé, les quelques études qui ont été menées sur la diversification des portefeuilles de prêts pair-à-pair révèlent une sous-diversification. En ce qui concerne les motivations des investisseurs en financement participatif, les études ont identifié deux formes : les motivations intrinsèque et extrinsèque. Ces études se rejoignent et montrent que la décision d'investissement en financement participatif est drainée, entre autres, par ces deux facteurs. Quant aux études sur les récits de demande de prêts, elles révèlent que leur contenu affecte les décisions d'investissement.

Ce rapide parcourt des recherches sur le comportement d'investissement des prêteurs dans le financement participatif, en particulier dans celui à but non lucratif, nous fait remarquer l'absence d'études ayant examiné s'il existe ou non des comportements propres à certains sous-groupes d'investisseurs. Pourtant, Lin et al. (2014; 2019) ont montré l'existence de sous-groupes de crowdfunders et de comportements d'investissement propres à chacun d'eux. Ces auteurs ont identifié dans leur première étude quatre types d'investisseurs (les actifs, les suiveurs de tendance, les altruistes et les spectateurs) et dans la seconde cinq (actifs silencieux, actifs vocaux, enthousiastes ciblés, suiveurs de tendances et chercheurs d'étoiles). Qu'ils soient silencieux ou vocaux, les actifs sont des investisseurs qui contribuent au financement d'un nombre important de projets et ont tendance à soutenir des projets précoces. Leur rôle sur les plateformes est donc capital et étudier leur choix d'investissement nous paraît primordial.

3. Objectifs et hypothèses de recherche

Cette étude poursuit trois objectifs principaux :

Objectif 1 : Étudier les stratégies de diversification adoptées par les multi-prêteurs

Il s'agit, au-travers de cet objectif, de mieux comprendre les choix de ces prêteurs. Pour cela, nous évaluons dans quelle mesure les portefeuilles de prêts sont diversifiés ou concentrés en considérant le genre des emprunteurs, leurs pays, leur secteur d'activité et s'ils se constituent en groupe ou non. Étant donné que les multi-prêteurs soutiennent plusieurs projets différents, nous nous attendons à ce que la composition des portefeuilles diffère d'un prêteur à un autre et que cette diversité se révèle dans le niveau de performance des portefeuilles. Nous formulons, par conséquent, les deux hypothèses suivantes :

Hypothèse 1-1 : Les multi-prêteurs adoptent différentes stratégies de diversification.

Hypothèse 1-2 : La performance des portefeuilles de prêts varie en fonction du niveau de diversification.

Objectif 2 : Examiner le lien entre les facteurs intrinsèques et extrinsèques présents dans les récits de demande de prêts et la durée des campagnes de financement des projets soutenus par les multi-prêteurs

Allison et al. (2015) ont montré que les prêteurs Kiva sont intrinsèquement motivés à soutenir des projets car ils réagissent positivement aux facteurs intrinsèques présents dans les récits de demande de prêts mais négativement à ceux extrinsèques. S'il est donc vrai que les emprunteurs qui mettent l'accent sur les facteurs intrinsèques dans leur récit attirent les prêteurs contrairement aux emprunteurs qui mettent l'accent sur les facteurs extrinsèques, alors cela pourrait être d'autant plus vrai pour les multi-prêteurs. Premièrement, parce qu'ils sont très actifs sur la plateforme et que les prêteurs qui sont très intrinsèquement motivés sont ceux qui participent le plus sur les plateformes de financement participatif. Deuxièmement, parce qu'ils viennent pour la plupart des États-Unis dont la culture est axée sur l'individualisme. Il est donc probable que ces prêteurs soient motivés par la satisfaction qui découle de l'aide apportée aux autres ou des actes philanthropiques (Moonhee et Gawon, 2017). Nous formulons ces deux hypothèses :

Hypothèse 2-1 : Un niveau élevé de langage intrinsèque est négativement associé à la durée des campagnes de financement des projets soutenus par des multi-prêteurs.

Hypothèse 2-2 : Un niveau élevé de langage extrinsèque est positivement associé à la durée des campagnes de financement des projets soutenus par des multi-prêteurs.

Objectif 3 : Évaluer l'effet d'être un groupe d'emprunteurs plutôt qu'un emprunteur seul sur la durée des campagnes de financement des projets sélectionnés par les multi-prêteurs

Dans la littérature, il existe une contradiction quant à la préférence des prêteurs pour les types d'emprunteurs. En effet, pour certains auteurs les prêteurs préfèrent les groupes mais pour d'autres auteurs leur préférence porte vers les emprunteurs seuls. Pham et Shen (2016) expliquent la préférence des prêteurs Kiva pour les groupes par le fait que ces prêteurs sont intrinsèquement motivés et perçoivent au-travers des groupes d'emprunteurs un moyen d'aider un plus grand nombre de personnes dans le besoin via un seul prêt. Nous croyons que ce raisonnement doit être plus vrai dans le cas particulier des multi-prêteurs. L'hypothèse est donc :

Hypothèse 3 : Un prêt de groupe nécessite moins de temps pour son financement qu'un prêt individuel.

4. Méthodologie

Nous consacrons ce chapitre à la méthodologie afin de décrire les méthodes utilisées pour atteindre nos trois objectifs. Nous le subdivisons donc en trois grandes parties (section 4.1, section 4.2 et section 4.3), une partie pour chaque objectif. Dans chaque partie, nous avons pris le soin de relever brièvement les méthodes utilisées dans les études précédentes, de justifier nos choix de méthodes avant de les présenter.

4.1. Diversification des portefeuilles de prêts

Cette section consacrée à la méthodologie sur la diversification regroupe le détail sur les méthodes que nous employons pour mesurer le niveau de diversification de chaque portefeuille (sous-section 4.1.1 et sous-section 4.1.2), pour créer des grappes de portefeuilles homogènes en fonction du niveau de diversification (sous-section 4.1.3) et pour évaluer comment la performance des portefeuilles est associée au niveau de diversification (sous-section 4.1.4).

4.1.1. Mesures de diversification des portefeuilles

Il existe dans la littérature plusieurs mesures pour estimer le niveau de diversification d'un portefeuille d'actions. Nous avons la perte relative du ratio de Sharpe (Calvet et al., 2009; Campbell et al., 2007), la variance normalisée du portefeuille (Goetzmann et al., 2005), le nombre total de titres (Van Horne et al., 1975) ou l'écart entre un portefeuille et le portefeuille du marché (Van Horne et al., 1975). Contrairement aux autres mesures, les deux dernières sont les plus adaptées et communément utilisées pour capter le niveau de diversification d'un portefeuille de prêts prosociaux (Lin et Bow, 2020; Mintjes, 2016). Le nombre total de prêts composant le portefeuille d'un crowdfunder est une mesure brute du niveau de diversification. Bien qu'elle soit simple et largement utilisée, cette méthode surestime le niveau de diversification (Blume et al., 1974). Pour ce qui est de l'écart d'un portefeuille donné par rapport au portefeuille du marché, cette mesure se définit comme suit :

$$DIV = \sum_{i=1}^{N} (w_i - w_m)^2 = \sum_{i=1}^{N} \left(w_i - \frac{1}{N_m} \right)^2 \quad (1)$$

Où N est le nombre de projets soutenus par le prêteur. N_m est le nombre de projets proposés sur la plateforme de Kiva. w_i est le poids du portefeuille qui est attribué à chaque projet i et w_m est le poids de chaque projet dans le portefeuille de la plateforme.

En considérant que la proportion de chaque titre dans le portefeuille du marché est très faible, Van Horne et al. (1975) supposent que $w_m = \frac{1}{N_m} \approx 0$. L'équation (1) devient donc :

$$DIV = \sum_{i=1}^{N} {w_i}^2 \qquad (2)$$

Les auteurs calculent donc la diversification comme la somme des carrés des proportions investies dans chaque titre, ce qui revient à utiliser l'indice Herfindahl-Hirschman (HHI). Ce dernier a été utilisé par Lin et Bow (2014 et 2019) pour examiner dans quelle mesure l'intérêt d'un crowdfunder est concentré sur un sujet, ou est diversifié sur plusieurs sujets. Nous choisissons donc d'utiliser l'IHH. Pour chaque projet disponible sur la plateforme, il y a quatre principales caractéristiques : le pays du porteur de projets, le secteur d'activité du projet, le genre du porteur de projets et le type du porteur de projets (individu, groupe ou autre). Chacune de ces caractéristiques est subdivisée en sous-caractéristiques qui donnent un aspect particulier à chaque projet de sorte qu'un projet avec une sous-caractéristique donnée peut être considéré comme un projet différent d'un projet avec une autre sous-caractéristique. Par exemple, pour le genre du porteur de projet, nous distinguons trois projets différents : les projets portés par des hommes, les projets portés par des femmes et les projets dont le genre du porteur de projet n'est pas précisé. En raisonnant de même, nous distinguons 98 projets différents pour les 98 pays dont sont issus les emprunteurs, 3 pour le type de l'emprunteur (individu, groupe ou autre) et 15 pour les secteurs d'activité.

Les prêteurs ont donc la possibilité de diversifier leur portefeuille non seulement au travers des quatre caractéristiques mais également au-travers des sous-caractéristiques de chaque caractéristique. Nous adaptons donc l'équation (2) pour mesurer le niveau de diversification des portefeuilles pour chacune des caractéristiques. On a :

$$DIV_j = \sum_{h=1}^{H} \left(\frac{n_{h,j}}{N}\right)^2$$

Où

$$H = \begin{cases} 3\ si\ j\ est\ le\ genre\ du\ porteur\ de\ projet \\ 3\ si\ j\ est\ le\ type\ de\ porteur\ de\ projet \\ 15\ si\ j\ est\ le\ secteur\ d'activité \\ 98\ si\ j\ le\ pays\ du\ porteur\ de\ projet \end{cases}$$

$n_{h,j}$ est le nombre de projets avec la sous caractéristique h de la caractéristique j que le prêteur a soutenus.

En suivant cette méthodologie, nous obtenons quatre séries de valeurs indiquant le niveau de diversification de chaque portefeuille pour les quatre caractéristiques.

4.1.2. Construction d'un indicateur synthétique

Afin d'apprécier le niveau global de diversification de chaque portefeuille de prêts, nous construisons un indicateur composite à partir des quatre mesures de diversification précédentes. Pour ce faire, nous suivons les étapes que voici :

Étape 1 : Réaliser le test de Bartlett et le test de Kaiser-Meyer-Olkin (KMO) :

Quel que soit l'objectif visé lorsqu'on envisage de réaliser une analyse en composantes principales (ACP), il faut réaliser des tests tels que le test de Bartlett et le test KMO pour déterminer si les données de l'échantillon conviennent à une telle analyse. Le test de Bartlett teste l'hypothèse selon laquelle la matrice de corrélation issue des données est égale à la matrice d'identité.

Soit $R = \left(r_{ij}\right)_{4x4}$ la matrice de corrélation et $|R|$ son déterminant. Sous l'hypothèse nulle, on a $|R| = 1$ et sous l'alternative $|R| = 0$. La statistique du test de Bartlett mesure à quel degré nous nous écartons de l'hypothèse nulle et est définie comme suit :

$$\chi^2 = -\left(192 - \frac{13}{6}\right)\ln|R|$$

Sous l'hypothèse nulle, la statistique suit une distribution de khi-deux avec un degré de liberté égal à 6. L'ACP sera considérée comme appropriée si le test de sphéricité de Bartlett est significatif ($p<0,05$) car cela impliquerait que les variables sont corrélées.

Avec le test de Kaiser-Meyer-Olkin, l'objectif est le même mais ici ce sont les corrélations partielles qui sont utilisées.

En désignant par $A = (a_{ij})$ la matrice de corrélations partielles. L'indice KMO global se calcule comme suit :

$$KMO = \frac{\sum_i \sum_{j \neq i} {r_{ij}}^2}{\sum_i \sum_{j \neq i} {r_{ij}}^2 + \sum_i \sum_{j \neq i} {a_{ij}}^2}$$

Pour chaque variable, l'indice est mesuré par :

$$MSA_j = \frac{\sum_{i \neq j} {r_{ij}}^2}{\sum_{i \neq j} {r_{ij}}^2 + \sum_{i \neq j} {a_{ij}}^2}$$

Des valeurs de KMO et des MSA supérieures à 0,8 sont jugées bonnes indiquant qu'une ACP est appropriée tandis que celles qui sont inférieures à 0,5 sont jugées inacceptables.

Étape 2 : Effectuer une analyse en composantes principales (ACP)

L'ACP est une technique de réduction de dimensions qui permet d'obtenir à partir des variables initiales de nouvelles variables qui sont des combinaisons linaires des premières. Elle permet, entre autres, d'explorer les relations entre les variables. La réalisation d'une ACP nous permettra donc d'obtenir les pondérations de chacune des mesures de diversification obtenues précédemment. L'avantage avec cette méthode est qu'elle nous évitera d'utiliser des pondérations subjectives, qui certainement seront erronées, pour construire l'indicateur synthétique (Bernile et al., 2018; Anago et Houngbeme, 2015; Zinsou, 2008). En effet, un indicateur global de diversification calculé comme une moyenne arithmétique des quatre variables suggère que toutes les variables ont la même importance, ce qui peut être erroné.

Pour choisir le nombre optimal d'axes à la suite de l'ACP, nous suivons trois des principaux critères qui sont habituellement utilisés : le pourcentage d'information restituée par les q premiers axes, le critère de Kaiser-Guttman et le critère du coude de Cattell. Selon le premier critère, il faut retenir les axes qui retournent entre 60% et 80%. Quant au deuxième, il stipule qu'il faudrait retenir tous les axes ayant des valeurs propres supérieures à l'unité. Enfin, Cattell recommande de représenter les pourcentages de variances

expliquées par chaque axe et de retenir les axes qui se situent avant le changement de pente de la représentation graphique.

En désignant par $w_k^{(s)}$ la coordonnée de la variable k sur l'axe factoriel s et λ_s la valeur de l'inertie expliquée par l'axe factoriel s, nous pouvons définir les K variables par :

$$DIV_k^{(s)} = w_k^{(s)} \times Axe_s + \mu_k$$

Où μ_k est la variance non expliquée par DIV_k.
Cela nous montre que chaque axe récupère une partie de l'information contenue dans chacune des K variables. Pour obtenir les axes, il faut donc déterminer le poids de l'information apportée par chaque variable dans la formation de chaque axe. Ledit poids pour la variable k sur l'axe s, peut s'obtenir par la formule suivante :

$$W_k^{(s)} = \frac{w_k^{(s)}}{\lambda_s}$$

Étape 3 : Calcul de l'indicateur

A cette étape, nous pouvons utiliser les poids déterminés précédemment pour définir chaque axe :

$$Axe_s = \sum_{k=1}^{K} W_k^{(s)} \times DIV_k$$

Les axes ainsi définis remplaceront les variables d'origine dans le calcul de l'indicateur global de diversification. Ce dernier sera alors la moyenne pondérée de ces axes. Le poids de chaque axe est égal au ratio entre la valeur de l'axe et la somme des valeurs propres des axes retenus, et représente la proportion de la variance totale prise en charge par l'axe considéré.

$$IG = \sum_{s=1}^{S} \frac{\lambda_s \times Axe_s}{\sum_{s=1}^{S} \lambda_s}$$

Étape 4 : Normaliser l'indicateur

Pour que l'indicateur ne prenne pas de valeurs négatives, nous ferons une transformation de la série en ajoutant à la valeur de l'indicateur pour chaque portefeuille la valeur absolue de la valeur minimale de la série.

$$IG_i^* = IG_i + |\min(IG)|$$

L'indicateur ainsi transformé variera entre 0 et 1. Les valeurs proches de 1 indiqueront une sous-diversification tandis que des valeurs proches de 0 indiqueront une importante diversification.

4.1.3. Méthodes de classification des portefeuilles

La classification a pour objectif de regrouper les données en grappes de manière à réduire la variance entre les éléments d'une même grappe et à maximiser la variance entre les grappes. Il s'agit donc d'une approche qui nous fournira une description assez riche et en même temps bien détaillée de nos données. En la réalisant, notre objectif est de distinguer les portefeuilles qui sont proches par leur niveau de diversification et ceux qui ne sauraient être assemblés car ils diffèrent significativement du fait de leur niveau de diversification.

On distingue deux principales méthodes de classification : hiérarchique et non hiérarchique. La classification hiérarchique consiste à construire, en plusieurs étapes, une structure arborescente en deux approches. La première, connue sous le nom d'agglomération, consiste à considérer chaque individu comme une grappe individuelle qui sera ensuite, à chaque étape du processus, fusionnée à la grappe individuelle la plus proche ou similaire. La seconde approche consiste à regrouper tous les individus dans une même grappe au départ et ensuite la diviser, à chaque étape, en plusieurs autres grappes.

Dans la classification non hiérarchique, encore appelée k-means, le nombre de grappes est défini et les centres sont calculés en avance (Hair et al., 1992). Ensuite, chaque individu est affecté à la grappe dont il est proche du centre. A chaque fois qu'un nouvel individu est ajouté à une grappe, le centre est recalculé. Ce processus est réitéré jusqu'à ce que l'on aboutisse à une solution optimale (Anderberg, 1973). Contrairement à la méthode

hiérarchique, cette méthode est moins susceptible d'être affectée par les observations aberrantes. De plus, grâce au processus d'itérations, cette méthode conduit à une solution qui optimise l'homogénéité au sein des grappes et l'hétérogénéité entre elles. Cependant, ces avantages reposent sur l'efficience des grappes initiales définies (Milligan, 1980).

En résumé, chacune des méthodes a ses avantages et ses désavantages. Il est donc recommandé de combiner les deux méthodes afin de tirer profit des avantages qu'offre chacune d'elles (Milligan, 1980 ; Punj et Stewart, 1983, Ketchen et Shook, 1996). Nous suivons donc ce procédé en réalisant premièrement une classification hiérarchique puis une classification non hiérarchique.

Pour décider du nombre optimal de classes avec la méthode hiérarchique, nous exploitons le dendrogramme. Dans le cas de la méthode non hiérarchique, nous utilisons le package NbClust disponible dans R. Il s'agit d'un algorithme qui combine 30 indices (par exemple l'indice de Calinski et Harabasz, l'indice de Duda, indice Gap…) qui sont généralement utilisés pour déterminer le nombre optimal de classes.

Enfin, nous calculons le coefficient de Kappa pour vérifier le niveau de concordance entre les solutions fournies par les deux méthodes de classification. Le coefficient se présente comme suit :

$$K = \frac{P_A - P_C}{1 - P_C}$$

Où P_A est la proportion d'accords par pairs et P_C est la proportion d'accords aléatoires. S'il y a un accord parfait entre les évaluateurs alors K serait égal à 1 sinon K serait inférieur ou égal à 0. Une valeur entre 0,41 et 0,60 est considérée comme un accord moyen, donc acceptable (Landis et Koch, 1977).

4.1.4. Régression des risques proportionnels de Cox

La régression de Cox ou régression des risques proportionnels examine et modélise la probabilité qu'un évènement précis se produise en fonction du temps et des variables explicatives (Spruance et al., 2004). Alors que la régression logistique exige une même date d'entrée dans l'étude pour tous les sujets étudiés afin d'éviter l'influence du moment d'entrée des sujets dans l'étude sur le résultat, la régression de Cox permet de prendre en

compte des dates d'entrée différentes. Dans notre cas, les sujets étudiés entament à différents moments leur campagne de financement.

De plus le modèle de Cox offre l'avantage de corriger les biais pouvant survenir en raison de la censure des données qui se produit lorsque l'évènement d'intérêt n'est pas observé pour certains individus durant la période d'étude. Les sujets que nous étudions ont en moyenne 30 jours pour mener leur campagne et se faire financer. Au-delà de ce délai, il n'est plus possible de les suivre pour observer la réaliser ou non de l'évènement d'intérêt qui est la réussite de la campagne de financement.

Le modèle de Cox nécessite à la fois une variable qualitative et une quantitative comme variables dépendantes. Ici, la variable quantitative est le nombre de jours qui sépare la date de début de la campagne de financement de chaque projet et sa date de fin, et la variable qualitative prend la valeur 1 si la campagne de financement a été un succès et 0 sinon. La forme fonctionnelle du modèle de Cox (1972), en fonction du temps t et des variables explicatives X se définit comme suit :

$$\lambda\left(t, X_1, X_2, \dots X_p\right) = \lambda_0(t) exp\left\{\sum_{k=1}^{p} \beta_k X_k\right\}$$

Avec $\lambda_0(t)$ le risque de base c'est-à-dire la probabilité que la campagne de financement réussisse lorsque toutes les variables explicatives sont nulles.

La méthode d'estimation est le maximum de vraisemblance mais compte tenu du fait que seules les campagnes encore en cours sont prises en compte dans l'estimation du modèle à chaque date t, on parle de vraisemblance partielle.

Sous l'hypothèse d'absence d'évènements simultanés et en notant T_i le nombre de jours qui s'est écoulé entre le début de la campagne du projet i et sa fin, la vraisemblance partielle du modèle se définit comme suit :

$$\mathcal{L}_i = \frac{\lambda(t, X_i)}{\sum_{T_j > T_i} \lambda(t, X_k)} = \frac{exp\left(X_i^{'}\beta\right)}{\sum_{T_j > T_i} exp\left(X_j^{'}\beta\right)}$$

Pour tout i appartenant à l'ensemble I des campagnes de financement non encore terminées, nous avons :

$$\mathcal{L} = \prod_{i \in I} \frac{exp\left(X_i^{'}\beta\right)}{\sum_{T_j > T_i} exp\left(X_j^{'}\beta\right)}$$

En y appliquant le logarithme népérien, nous obtenons :

$$\ln \mathcal{L} = \sum_{i \in I} \left(X_i'\beta - \ln \left(\sum_{T_j > T_i} exp\left(X_j'\beta\right) \right) \right)$$

En maximisant cette expression nous obtenons les coefficients β_k qui représente l'effet des variables explicatives correspondantes sur la probabilité de réussite du financement des projets. Si β_k est nul alors la k-ième variable explicative n'a pas d'influence sur la probabilité de survenance de l'événement considéré. Toutes choses étant égales, si β_k est positif alors des valeurs élevées de la k-ième variable explicative sont associées à une probabilité de réalisation de l'évènement considéré plus élevée. Cependant si β_k est négatif, des valeurs élevées de la k-ième variable explicative sont associées à une probabilité de réalisation de l'évènement considéré plus faible.

Dans le cas où il y a des projets qui ont achevé leur campagne de financement à la même date, trois méthodes existent pour adapter la vraisemblance partielle et prendre en compte cette situation. En utilisant la méthode exacte l'on considère que les dates de fin de campagne sont bien différentes les unes des autres et que les égalités observées ne sont qu'apparentes et ne résultent que d'un manque de précision dans les mesures. Dans la méthode de Breslow (1974), on suppose que tous les événements qui se sont réalisés simultanément ont la même probabilité de réalisation que le premier d'entre eux. En revanche, dans la méthode d'Efron (1977), il est question de calculer une probabilité moyenne pour tous les événements survenus simultanément ce qui rend cette méthode plus performante que celle de Breslow. Elle est, par conséquent, la méthode disponible par défaut dans R et nous l'appliquons dans cette analyse.

La régression de Cox repose sur une hypothèse majeure qui est celle des risques proportionnels c'est-à-dire que pour deux projets avec des caractéristiques différentes la probabilité que l'un soit totalement financé est proportionnelle à celle de l'autre.

Pour tout t et toute paire de projets {i, j}, le rapport des probabilités $\lambda_i(t) = \lambda(t, X_i)$ et $\lambda_j(t) = \lambda\big(t, X_j\big)$ doit donc reste indépendant du temps. Soit

$$\frac{\lambda(t, X_i)}{\lambda\big(t, X_j\big)} = \frac{\lambda_0(t) exp(X_i'\beta)}{\lambda_0(t) exp\big(X_j'\beta\big)}$$

$$= \frac{exp(X_i'\beta)}{exp\big(X_j'\beta\big)}$$

$$= \exp\big((X_i' - X_j')\beta\big) = \text{constante}$$

Cependant, pour la plupart des études cette hypothèse n'est souvent pas vérifiée (Stensrud et Hernán, 2020; Schemper, 1992). Cela se produit surtout lorsqu'on travaille sur des données de grande taille. Rulli et al. (2018) montrent que les tests de non-proportionnalité peuvent manquer de puissance pour les échantillons de petite taille mais deviennent trop sensibles à de légers écarts pour les échantillons de grande taille. Une non-proportionnalité significative peut, par conséquent, être sans incident majeur sur les résultats et leur interprétation pour les données de grande taille (Therneau et Grambsch, 2000). La décision d'ignorer une non-proportionnalité significative dépend donc de l'étude menée. Dans notre cas, les anciennes études (Allison et al., 2013 et 2015) ne font mention d'un quelconque incident pouvant résulter de la violation de cette hypothèse. Compte tenu de la taille de nos données, nous nous attendons à ce que l'hypothèse de risques proportionnels ne soit pas vérifiée mais que les résultats ne soient pas pour autant invalides.

4.2. Récits de demande de prêt et durée des campagnes de financement

Dans la recherche sur l'importance des récits de demande de prêt pour les campagnes de collecte de fonds en financement participatif, l'approche statistique communément adoptée est une régression multiple. Grâce à cette méthode, il est possible de déterminer les associations spécifiques de certaines variables avec la variable dépendante, tout en contrôlant simultanément celles que d'autres variables pourraient également avoir. Cependant, le choix du type de régression qu'il convient d'appliquer dépend de la nature de la variable dépendante retenue dans l'analyse. Elle peut être une variable binaire mesurant le succès ou non des campagnes de financement ou une variable quantitative mesurant le nombre de jours ou de minutes nécessaire pour qu'un projet soit totalement financé. Dans le cas où la variable binaire est utilisée, on retrouve habituellement la régression logistique ou probit comme c'est le cas dans l'étude de Larrimore et al. (2011) et celle de Parhankangas et Renko (2017). Moss et al. (2015) ont, quant à eux, ont opté pour le modèle à risque proportionnel de Cox car il tient compte du biais que pourrait engendrer toute forme de censure dans les données.

L'intervention des IMF dans le processus de financement des prêts en tant que partenaires de terrain de Kiva crée une certaine forme d'hiérarchisation dans les données, lorsque des variables telles que la cote de risque des partenaires de terrain est utilisée dans l'analyse. Moss et al. (2017) ont choisi donc d'utiliser un modèle linéaire hiérarchique à trois niveaux afin de prendre en compte la non-indépendance entre les observations due à la présence de variables liées aux partenaires de terrain et aux secteurs d'activité des emprunteurs (Peterson et al., 2012).

Allison et al. (2013 et 2015) ont appliqué la régression des moindres carrés ordinaires et le modèle de Cox afin d'estimer comment les indices linguistiques sont associés à la durée des campagnes de financement, tout en contrôlant plusieurs autres facteurs.

L'objectif de notre analyse est d'examiner les mêmes associations que Allison et al. (2015) mais pour le groupe des « multi-prêteurs ». Nous choisissons, par conséquent, d'appliquer comme eux, la régression des moindres carrés ordinaires et un modèle de risque proportionnel de Cox pour vérifier à quel point les résultats sont sensibles à une modification dans la méthodologie.

En appliquant une régression MCO, nous supposons qu'il existe une relation linéaire entre chacune des variables indépendantes et la variable dépendante ainsi qu'entre chacune des variables de contrôle et la variable dépendante. Pour satisfaire à l'hypothèse de normalité des résidus que requiert la régression MCO, plutôt que d'utiliser directement la variable *Duree* dans le modèle, nous y appliquons le logarithme népérien.

Soit Y_i la durée qu'il faut pour que le projet i soit totalement financé, D_i les variables indépendantes et C_i les variables de contrôle. Le modèle linéaire se présente comme suit :

$$\log(Y_i) = \beta_0 + \beta_1 D_i + \beta_2 C_i + \varepsilon_i \qquad (3)$$

Avec ε_i le terme d'erreur. Le test de Breusch-Pagan indique la présence d'hétéroscédasticité ($p < 2e\text{-}16$) dans les données. Nous recourons donc à la procédure de Eicker-Huber-White pour obtenir des erreurs standards consistantes.

4.3. Prêts de groupe versus prêts individuels

Notre objectif est d'étudier l'effet d'être un groupe d'emprunteurs plutôt qu'un emprunteur seul sur la durée des campagnes de financement. Il s'agit donc d'une étude d'inférence causale. La seule étude, à notre connaissance, ayant réalisé une telle analyse pour le financement participatif basé sur la dette est celle de Pham et Shen (2016). En effet, les autres études ayant traité des prêts de groupe en financement participatif ont surtout cherché à voir comment la décision des emprunteurs de se constituer en groupe ou de demeurer seuls se trouve être associée à la durée des campagnes de financement (Galak et al., 2011; Berns et al., 2020).

Les emprunteurs en groupe sont donc, dans l'analyse causale que nous menons, considérés comme les traités et les emprunteurs seuls comme les témoins. Étant donné que nous utilisons des données observationnelles, nous recourrons au modèle causal de Rubin (Rubin, 1974) dont la formalisation et les hypothèses sont présentées dans la sous-section 4.3.1.

Le fait de chercher à prendre en compte les récits de demande de prêts nous amène à faire face à un défi colossal à cause de la nature non structurée des données textuelles. Dans la sous-section 4.3.3 nous présentons donc les méthodes d'inférence causale adaptées au cas de données observationnelles importantes que nous avons choisies d'utiliser, après avoir relevé à la sous-section 4.3.2 les méthodes développées dans la littérature à cet effet. Enfin, nous présentons dans la sous-section 4.3.4 les modèles de régression des résultats et de score de propension.

4.3.1. Le cadre des résultats potentiels

Le modèle causal de Rubin offre un cadre pour l'estimation de l'effet d'un traitement à partir de données observationnelles. Pour chacun des sujets, il y a deux résultats possibles : un résultat pour le cas où le sujet considéré appartient au groupe traité ($Y(1)$) et un autre correspondant au cas où il fait partie du groupe témoin ($Y(0)$). Puisque dans les données réelles nous avons qu'un seul de ces deux résultats pour chaque sujet, l'objectif dans ce modèle est de déduire le résultat non observable à partir de celui observable afin de déterminer l'effet recherché comme l'écart entre ces deux résultats. En considérant tous les sujets, cela revient à déterminer l'effet moyen du traitement comme suit :

$$\tau = \mathbb{E}[Y(1)] - \mathbb{E}[Y(0)]$$

Avec X l'ensemble des covariables, Z la variable binaire qui prend la valeur 1 si l'emprunteur est un groupe et 0 sinon et τ l'effet moyen de Z sur la variable Y qui mesure en jour la durée des campagnes de financement.

Comme $Y(1)$ et $Y(0)$ correspondent respectivement au cas où Z=1 et Z=0, cela suggère que le résultat observable Y^{obs} doit satisfaire à la relation suivante :

$$Y^{obs} = Z * Y(1) + (1 - Z) * Y(0)$$

Pour déterminer l'effet moyen d'intérêt au moyen du modèle des résultats potentiels, trois hypothèses sont nécessaires (Zhichao et al., 2020; Wijayatunga, 2014; Ding et Li, 2018).

La première est appelée hypothèse de valeur de traitement unitaire stable (HVTS) qui, en fait, regroupe deux hypothèses : une qui suppose que les résultats, pour toute observation, ne varient pas avec les traitements des autres observations, et la seconde suppose qu'il n'existe pas de formes ou de versions différentes de traitement qui conduisent à des résultats différents. L'hypothèse HVTS sera vérifiée si, pour chaque projet, le ou les porteurs du projet font face au même nombre de niveaux de traitement et que le résultat potentiel est indépendant du traitement subi par les autres projets.

Dans le cadre de cette étude, la variable de traitement a exactement deux niveaux de traitement (groupe ou individu) auxquels tous les porteurs de projet font face. La première partie de l'hypothèse HVTS est donc vérifiée. La variable de résultat est la durée de la campagne de financement et dépend principalement des caractéristiques propres à chaque projet. De plus, la seule contrainte à laquelle les prêteurs Kiva sont soumis en ce qui concerne le montant à accorder est un minimum de 25 $ US. Ce montant est trop petit pour empêcher un prêteur de soutenir le nombre de projet qu'il souhaite. Enfin, même si la présence des partenaires de terrain de Kiva pourrait porter à croire à une potentielle interférence, cela n'est pas vrai car le traitement appliqué à chaque projet dépend étroitement des caractéristiques qui lui sont propres.

La deuxième hypothèse est l'hypothèse de non-confusion ou d'ignobilité :

$$Y(1), Y(0) \perp Z/X$$

Selon cette hypothèse, conditionnellement aux caractéristiques X, les résultats $Y(1)\ et\ Y(0)$ sont indépendants du traitement Z. Cette hypothèse suppose qu'il n'existe aucune variable non observée ou non observable qui pourrait affecter simultanément Z et Y. En d'autres termes, si deux projets ont les mêmes caractéristiques observables X alors leurs résultats potentiels doivent être les mêmes indépendamment du traitement subit. En assumant cette hypothèse, on peut assigner la différence $Y(1) - Y(0)$ à Z.

Dans le cas des emprunteurs Kiva, leur décision de se constituer en groupe ou de rester seul peut être considérée, dans une certaine mesure comme exogène, car l'importance accordée aux prêts de groupe repose sur la connaissance que les emprunteurs ont de leurs pairs. Cela suggère que les emprunteurs qui se mettent en groupe sont généralement des connaissances et qui le font plus par commodité que pas soucis de réduire la durée de leur campagne (Pham et Shen, 2016). De plus, étant donné que l'idée des prêts de groupe date des années 70, s'il s'avère réellement que les emprunteurs ont connaissance de l'avantage potentiel qu'il y a à se constituer en groupe alors nous devrions constater un nombre prépondérant de prêts de groupe parmi les demandes de prêts, ce qui n'est pas le cas. Nous pouvons donc considérer que si elle existe, cette endogénéité est bien faible et peut être négligée.

La troisième et dernière hypothèse est celle de la positivité, encore appelée hypothèse de chevauchement. Elle soutient que conditionnellement à X, l'affectation au traitement n'est pas déterministe autrement dit que la relation suivante est vraie avec une probabilité de 1 :

$$0 < \mathrm{P}(Z = 1/\mathrm{X}) < 1$$

Cette hypothèse suggère que chaque sujet à une probabilité non nulle d'être traité et d'être non traité, ce qui assure d'avoir des observations dans le groupe des traités et dans le groupe des témoins. Admettons par exemple, que tous les projets provenant d'Afrique de l'Ouest sont portés par des emprunteurs seuls, il serait alors inutile voire impossible d'étudier l'effet de la formation de groupe sur la durée des campagnes de financement pour ces emprunteurs. Cette hypothèse garantie donc une certaine variabilité rendant ainsi possible l'estimation de l'effet causal.

Ces trois hypothèses sont désignées sous l'appellation d'hypothèse d'ignorabilité forte ou d'affectation de traitement fortement ignorable (Rosenbaum et Rubin, 1983; Markozannes et al., 2021). Grâce à elles, il est possible d'identifier les effets de traitement à partir de données observationnelles.

4.3.2. Inférence causale en présence de données non randomisées de grande dimension

En matière d'inférence causale, il est primordial de contrôler pour l'ensemble des facteurs de confusion, c'est-à-dire des variables qui influencent à la fois le traitement et le résultat. La correction des inférences pour ces facteurs est d'autant plus importante lorsqu'on utilise des données observationnelles car en absence de randomisation les réponses observées ne sont pas nécessairement équilibrées entre le groupe des traités et celui des non traités, ce qui peut engendrer un biais de confusion.

De récentes études en sciences sociales ont montré que les données textuelles peuvent servir de substituts à ces facteurs de confusion (Sridhar et Getoor, 2019; Sridhar et al., 2018) ou carrément constituer elles-mêmes des facteurs de confusions non mesurées dans les données structurées (Zou, 2006; Keith et al., 2020; Weld et al., 2021). Pour ces raisons et surtout à cause de l'importance que revêt les récits de demande de prêts en recherche entrepreneuriale (Cornelissen et Clarke, 2010 ; Martens et al., 2007 ; Zott et Huy, 2007), les auteurs Pham et Shen (2016) ont jugé bon de les introduire dans leur analyse. Cependant, ces données sont de nature non structurée et de haute dimension, ce qui rend leur traitement avec les méthodes standards très difficile voire inefficace. Cette problématique a fait l'objet d'un grand intérêt au cours des dernières années conduisant à la proposition de nombreux estimateurs potentiels permettant d'estimer les effets causaux lorsque le nombre de facteurs de confusion potentiels est important. A titre indicatif, nous pouvons mentionner le lasso adaptatif (Wilson et Reich, 2014; Shortreed et Ertefaie, 2017), l'estimateur de vraisemblance pénalisé (Ertefaie et al., 2018), l'estimateur bayésien (Antonelli et al., 2019), l'estimateur de Hahn et al. (2016), l'estimateur de double sélection (Belloni et al., 2014), l'estimateur doublement robuste (Robins et al., 1994) ou le maximum de vraisemblance ciblé (Van Der Laan et Rubin, 2006).

Le fait qu'il existe une pléthore d'estimateurs potentiels pour estimer les effets causaux dans un contexte de haute dimension, crée un problème de choix. Athey et al (2017) ont identifié l'estimateur de double sélection, celui doublement robuste et l'estimateur ciblé du maximum de vraisemblance comme les estimateurs les plus performants car ils permettent de corriger à la fois les biais dus à l'association entre les variables de prétraitement et les résultats potentiels et les biais provenant de l'association entre les variables de

prétraitement et l'affectation du traitement. La méthode de double sélection assure une inférence uniformément valide pour les effets causaux en cas de données à dimensions élevées (Belloni et al., 2014). L'estimateur doublement robuste, quant à lui, accroit les chances d'obtenir des effets causaux valides car il possède la propriété de double robustesse. En inférence causale, la propriété de double robustesse est très prisée car il suffit que le modèle d'estimation du résultat ou celui de l'affectation au traitement soit correctement spécifié pour que l'estimateur soit cohérent. Tout comme l'estimateur doublement robuste, qui n'en est qu'un cas particulier, l'estimateur ciblé du maximum de vraisemblance possède également cette propriété de double robustesse.

Selon Antonelli et Cefalù (2020), il n'est pas souvent clair si ou quand un estimateur est préférable à un autre puisque chaque estimateur a des caractéristiques différentes, ce qui rend subjectif le choix de l'estimateur à appliquer. Ces auteurs proposent donc un estimateur synthétique qui prend la moyenne des estimateurs candidats. Ils montrent que cet estimateur moyen est un bon compromis entre un très mauvais estimateur et un très bon estimateur car il évite à l'analyste d'opter pour un estimateur très mauvais et fonctionne aussi bien que le meilleur des estimateurs candidats.

Pour leur étude, Pham et Shen (2016) ont utilisé l'estimateur de double sélection, l'estimateur doublement robuste et l'estimateur ciblé du maximum de vraisemblance auxquels ils ont ajouté une régression régularisée ElasticNet afin de montrer la sous performance des modèles économiques standards dans le cas où l'on dispose de données à haute dimension. Les auteurs justifient ces trois choix en soulignant que l'estimateur de double sélection est facile à implémenter, moins coûteux en temps de calcul et fournit une estimation acceptable dans de nombreux cas. Quant à l'estimateur doublement robuste et l'estimateur ciblé par maximum de vraisemblance, leur propriété de double robuste font d'eux des estimateurs très intéressants. À l'issu de leur étude, les auteurs ont tiré la conclusion selon laquelle les prêts de groupe se financent 3,3 jours plus vite, en se basant sur l'estimateur doublement robuste et l'estimateur ciblé par maximum de vraisemblance implémentés avec la méthode d'estimation du réseau de neurones récurrent à longue mémoire à court terme. Ce choix repose à la fois sur la propriété de double robustesse et l'estimation du score de propension presque parfaite obtenue avec le réseau de neurones récurrent à longue mémoire à court terme.

Étant donné notre objectif et le fait que nos données proviennent de la même source que Pham et Shen (2016), nous choisissons d'adopter la même méthodologie qu'eux. Cela nous permettra également d'aisément comparer nos résultats qui porteront sur un sous-groupe particulier de prêteurs aux leurs qui portent sur tous les prêteurs Kiva.

Toutefois, nous apportons deux changements à la méthodologie de ces auteurs. En effet, nous utilisons un algorithme génétique pour déterminer la structure des réseaux de neurones et nous adoptons la méthode de Antonelli et Cefalù (2020) pour obtenir un estimateur moyen.

4.3.3. Méthodes d'estimation de l'effet causal

Dans cette sous-section, nous présentons en détail les méthodes utilisées pour estimer l'effet moyen des prêts de groupe sur la durée des campagnes de financement : la méthode naïve (sous-section 4.3.3.1), la régression pénalisée ElasticNet (sous-section 4.3.3.2), la méthode de la double sélection (sous-section 4.3.3.3), la méthode doublement robuste (sous-section 4.3.3.4), l'estimateur ciblé du maximum de vraisemblance (sous-section 4.3.3.5) et l'estimateur moyen (sous-section 4.3.3.6).

Pour la première méthode, aucun ajustement n'est effectué pour les variables confondantes. Cet estimateur n'estime donc pas sans biais l'effet causal d'intérêt. Quant aux deux méthodes (régression pénalisée ElasticNet et méthode de la double sélection) qui suivent celle-là, l'ajustement se fait en réalisant un modèle de régression des résultats dans lequel les variables confondantes seront considérées comme des variables de contrôle.

Dans le cas particulier de l'estimateur doublement robuste et de l'estimateur ciblé du maximum de vraisemblance, on fait deux types d'ajustements en réalisant deux modèles : un modèle de régression des résultats et un modèle de score de propension. Dans le modèle de régression des résultats, les variables confondantes seront considérées comme des variables de contrôle. Pour le second modèle, les variables de confusion sont utilisées dans le calcul du score de propension qui mesure la probabilité d'être traité sachant les covariables observées avant traitement (Rosenbaum et Rubin, 1983; Li, 2011; Guo et Fraser, 2014). Dans notre cas, il s'agit de la probabilité d'être un groupe d'emprunteurs plutôt qu'un emprunteur seul sachant les caractéristiques observées.

4.3.3.1. La méthode naïve

Cette méthode d'estimation de l'effet moyen de traitement consiste une simple différence de moyennes entre les traités et les témoins, soit :

$$\hat{\mathbb{T}}_{naïve} = \bar{Y}^{(1)} - \bar{Y}^{(0)}$$

Avec $\bar{Y}^{(1)}$ la moyenne des durées de campagne de financement observées pour les groupes d'emprunteurs qui représentent les traités et $\bar{Y}^{(0)}$ la moyenne des durées de campagnes de financement observées pour les emprunteurs seuls qui représentent les témoins.

4.3.3.2. Régression régularisée

La régression pénalisée avec la pénalité ElasticNet est une méthode de régression régularisée qui combine linéairement les pénalités L1 et L2 des méthodes lasso et crête. La régularisation, en elle-même, est une approche qui permet d'éviter le surajustement, surtout lorsque les données varient beaucoup. Par conséquent, elle permet de trouver un compromis entre le bais et la variance c'est-à-dire accepter une légère augmentation du biais pour bénéficier d'une réduction de la variance. Dans la pratique, il s'agit d'imposer des contraintes sur les coefficients estimés dans la régression.

Le modèle de régression peut s'écrire comme suit :

$$Y = X\beta + \varepsilon$$

Dans le cas d'une régression classique, pour estimer les paramètres de ce modèle, il suffit de minimiser l'espérance de l'écart au carré entre la prédiction $\hat{Y}$ et la vraie valeur Y. En revanche, lorsque la régulation est appliquée, il est plutôt question d'une minimisation sous contrainte comme ci-dessous :

$$\min_{\beta_1,\dots,\beta_p} \sum_{i=1}^{n}\left(Y_i - \sum_{j=1}^{p}\beta_j X_{ij}\right)^2$$

$$\text{S.C} \qquad \alpha\sum_{j=1}^{p}|\beta_j| + (1-\alpha)\sum_{j=1}^{p}\beta_j^2 \leq \tau$$

$$\min_{\beta_1,\dots,\beta_p} \sum_{i=1}^{n}\left(Y_i - \sum_{j=1}^{p}\beta_j X_{ij}\right)^2 + \lambda\left[\alpha\sum_{j=1}^{p}|\beta_j| + (1-\alpha)\sum_{j=1}^{p}\beta_j^2\right]$$

$$\min_{\beta_1,\ldots,\beta_p} \quad \sum_{i=1}^{n}\left(Y_i - \sum_{j=1}^{p}\beta_j X_{ij}\right)^2 + \lambda_1 \sum_{j=1}^{p}|\beta_j| + \lambda_2 \sum_{j=1}^{p}\beta_j^2$$

Par conséquent

$$\hat{\beta} = arg\min_{\beta}|Y - X\beta|^2 + \lambda_2|\beta|^2 + \lambda_1|\beta|_1$$

Où λ_1 et λ_2 sont des paramètres positifs à fixer. Si $\lambda_1 = 0$ et $\lambda_2 > 0$ alors il s'agit d'une régression de crête, Si $\lambda_1 > 0$ et $\lambda_2 = 0$ alors il s'agit d'une régression lasso et Si $\lambda_1 = 0$ et $\lambda_2 = 0$ alors il s'agit d'une régression OLS. Pour déterminer, les valeurs de λ_1 et λ_2 qu'il convient d'utiliser, nous avons utilisé la fonction train du package caret disponible dans R.

Nous appliquons ce modèle aux données sans les récits de demande de prêts d'une part et avec les récits d'autre part. En excluant les récits de demande de prêts, nous obtenons un modèle avec 17 variables explicatives mais en les y incluant on en a 117.

Dans les deux cas et séparément pour les emprunteurs constitués en groupe et les emprunteurs seuls, nous estimons dans un premier temps les coefficients β et par la suite nous prédisons les durées de campagne de financement. Ce qui nous permet d'avoir :

$$\hat{Y}^{(1)} = X^*\hat{\beta}^{(1)} \text{ et } \hat{Y}^{(0)} = X^*\hat{\beta}^{(0)}.$$

Ensuite, nous estimons l'effet recherché en calculant la moyenne des écarts entre les durées estimées comme suit :

$$\hat{\mathrm{T}}_{enet} = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{Y}^{(1)}{}_i - \hat{Y}^{(0)}{}_i\right)$$

L'erreur standard estimée est donnée par l'expression suivante :

$$\sqrt{Var^{(1)} + Var^{(0)}}$$

Où $Var^{(1)} = \frac{\mathrm{var}\left(\mathrm{Y_i} - \hat{\mathrm{Y}}^{(1)}{}_\mathrm{i} | \mathrm{i} : \mathrm{Z_i}=1\right)}{\mathrm{n}^{(1)}}$ et $Var^{(0)} = \frac{var\left(Y_i - \hat{Y}^{(0)}{}_i | i : Z_i=0\right)}{n^{(0)}}$

4.3.3.3. Estimateur de la double sélection

Proposée par Belloni et al. (2013), cette approche se déroule en trois étapes principales. La première consiste à effectuer une régression lasso pour sélectionner les variables qui expliquent la durée des campagnes de financement des projets de groupe et celle des projets individuels. Une seconde régression lasso est effectuée à la deuxième étape pour identifier les variables qui déterminent la décision des emprunteurs de présenter une demande de prêt en se constituant en groupe ou en étant seul. Une fois ces deux ensembles de variables identifiés, nous créons un nouvel ensemble qui constitue l'union des deux. A la dernière étape, nous effectuons une régression OLS pour les groupes d'emprunteurs et les emprunteurs seuls séparément en utilisant l'ensemble de variables obtenu après l'union. L'effet moyen du traitement estimé est obtenu en calculant la moyenne des différences entre la durée des campagnes de financement estimée pour les groupes et celle estimée pour les emprunteurs seuls. Les formules pour l'effet estimé et l'erreur standard estimée sont donc identiques à celles présentées pour la régression pénalisée ElasticNet.

4.3.3.4. Estimateur doublement robuste

Comme son nom l'indique, l'estimateur que l'on obtient à partir de cette approche possède la propriété de la double robustesse. En d'autres mots, la robustesse de l'estimateur (convergence et normalité asymptotique) est assurée dès lors que le modèle de régression des résultats ou celui du score de propension est bien spécifié. Cette propriété est capitale, en particulier lorsque l'on dispose d'un grand nombre de variables explicatives comme c'est le cas dans cette analyse car il est quasiment impossible d'être sûr que le modèle d'estimation du résultat et celui du score de propension sont tous deux bien spécifiés (Robins et Rotnitzky, 1995; Robins et al., 1994; Scharfstein et al., 1999).

Nous suivons ici également trois grandes étapes pour estimer l'effet moyen du traitement. D'abord, nous estimons les coefficients des modèles présentés à la sous-section 4.3.4 et à partir des coefficients estimés nous prédisons les durées des campagnes ou les scores de propensions, selon le cas.

Enfin, nous déduisons l'effet estimé recherché par :

$$\widehat{\mathrm{T}}_{EDR} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{Z_i\left(Y_i - \hat{Y}_i^{(1)}\right)}{\hat{e}_i} + \hat{Y}_i^{(1)}\right) - \frac{1}{n}\sum_{i=1}^{n}\left(\frac{(1-Z_i)\left(Y_i - \hat{Y}_i^{(0)}\right)}{1-\hat{e}_i} + \hat{Y}_i^{(0)}\right)$$

Où

$Z_i = 1$ indique que le porteur du projet i est un groupe et $Z_i = 0$ qu'il s'agit d'un seul emprunteur

Y_i est la valeur observée de la durée des campagnes de financement pour chaque projet i

$\hat{Y}_i^{(1)}$ est la valeur prédite de $\mathrm{E}(Y|X)$ pour $Z_i = 1$

$\hat{Y}_i^{(0)}$ est la valeur prédite de $\mathrm{E}(Y|X)$ pour $Z_i = 0$

$\hat{e}_i$ représente, pour chaque i, la valeur prédite du score de propension $e(X) = P(Z = 1|X)$, $0 < e(X) < 1$. Par score de propension, nous entendons la probabilité qu'un projet soit un projet de groupe étant donné les caractéristiques observées X du projet.

Lunceford et Davidian (2004) ont montré que l'estimation de la variance par la méthode sandwich de manière empirique est plus stable. Nous utilisons donc ce même procédé pour estimer l'erreur standard $\widehat{\mathrm{T}}_{EDR}$. Nous calculons pour chaque observation i :

$$IC_i = \left(\frac{(1-Z_i)(Y_i - \hat{Y}_i^{(1)})}{\hat{e}_i} + \hat{Y}_i^{(1)}\right) - \left(\frac{(1-Z_i)\left(Y_i - \hat{Y}_i^{(0)}\right)}{1-\hat{e}_i} + \hat{Y}_i^{(0)}\right) - \widehat{\mathrm{T}}_{EDR}$$

Puis

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n} {IC_i}^2$$

Et l'erreur standard recherchée est $\frac{\sigma}{\sqrt{n}}$.

4.3.3.5. Estimateur ciblé du maximum de vraisemblance

Cette méthode est une généralisation de la méthode doublement robuste décrite plus haut. La démarche reste donc la même et nous obtenons :

$$\widehat{\Psi}_{ECMV} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{A_i\left(Y_i - \hat{Q}_i^{a=1}\right)}{\hat{e}_i} + \hat{Q}_i^{a=1}\right) - \frac{1}{n}\sum_{i=1}^{n}\left(\frac{(1-A_i)\left(Y_i - \hat{Q}_i^{a=0}\right)}{1-\hat{e}_i} + \hat{Q}_i^{a=0}\right)$$

Avec $\hat{Q}_i^a = \hat{Y}_i^a - \hat{\varepsilon}H_i$, $H_i = \frac{A_i}{\hat{e}_i} - \frac{1-A_i}{1-\hat{e}_i}$ et $\hat{\varepsilon} = \frac{H_i\left(Y_i - \hat{Y}_i^a\right)}{\sum_{i=1}^{n} H_i}$ où $a = \{0,1\}$

L'expression de l'erreur standard estimée reste identique à celle de l'EDR. En plus de disposer de la propriété de double robustesse, cette méthode met à jour l'estimation initiale de $\hat{Y}_i^a$ afin que l'estimateur ciblé soit déterminé ce qui la rend plus précise que l'EDR (Van Der Laan et Rubin, 2006; Maya Petersen,2019).

4.3.3.6. L'estimateur moyen

Cet estimateur a été proposé par Antonelli et Cefalu pour répondre à la difficulté que crée l'existence de plusieurs estimateurs potentiels dans le contexte de données de haute dimension. Les auteurs soulèvent le fait qu'il n'est pas évident pour un analyste d'identifier avec exactitude l'estimateur qui convient le mieux à ses données. Il montre que cet estimateur synthétique est robuste et présente une performance supérieure à celle du pire estimateur et une performance aussi bonne que les meilleurs estimateurs. Dans le cas où il y a des contradictions entre les estimateurs, les auteurs montrent que l'estimateur synthétique se range du côté de la majorité, dans la plupart des cas.

Soit M le nombre d'estimateurs candidats, $\hat{\theta}_m$ avec m= 1, 2, …, M les effets moyens de traitement estimés pour chaque estimateur individuel. L'estimateur moyen se définit comme suit :

$$\hat{\theta}_{em} = \sum_{m}^{M} w * \hat{\theta}_m$$

Avec $w = \frac{1}{M}$ est le poids attribué à chaque estimateur individuel.

La variance estimée est donnée par :

$$\hat{\sigma}_{em}^2 = w^2 \sum_{i=1}^{M} \sum_{j=1}^{M} \hat{\sigma}_i \hat{\sigma}_j$$

Des dix estimateurs disponibles par défaut dans R pour cet estimateur moyen, nous ne retenons que l'estimateur du double post-sélection qui correspond à l'estimateur de double sélection présenté plus haut. Nous y ajoutons l'estimation obtenue avec la régression pénalisée, celle obtenue avec la méthode doublement robuste et celle obtenue avec l'estimateur ciblé du maximum de vraisemblance. Notre objectif, en procédant de la sorte, est de parvenir à une valeur unique pour l'effet moyen des prêts de groupe sur la durée des campagnes de financement car comme Pham et Shen (2016) nous nous attendons à obtenir des valeurs estimées différentes pour chaque méthode même si elles sont de même signe.
Nous faisons cette estimation avec la fonction AveragingCausalHD disponible dans R.

4.3.4. Modèles de régression des résultats et du score de propension

Pour la méthode doublement robuste et celle de l'estimateur ciblé du maximum de vraisemblance, il nous faut réaliser en amont un modèle de régression des résultats de traitement et un modèle de score de propension. Pour cela, nous recourrons à quatre approches : la régression pénalisée ElasticNet (sous-section 4.3.4.1), la méthode des forêts aléatoires (sous-section 4.3.4.2), un perceptron multicouche (sous-section 4.3.4.3) et un réseau de neurones récurrent profond avec une mémoire longue à court terme (sous-section 4.3.4.4).

Les deux dernières sont des méthodes d'apprentissage automatique. Le recours à cette technique pour traiter les données textuelles est assez répandu en économie et en sciences sociales. Par exemple, Gentzkow et al. (2015) ont recouru à des outils de traitement de texte pour traiter les discours des membres du Congrès afin d'évaluer dans quelle mesure un observateur peut déduire le parti d'un membre du Congrès à partir d'une quantité fixe de discours. Kang et al. (2013), quant à eux, ont obtenu une précision de 82% en utilisant les critiques rédigées sur Yelp, qui sont des données textuelles, pour prédire les niveaux d'hygiène des restaurants. Grâce à un réseau neuronal convolutif, Jean et al. (2016) ont pu obtenir des estimations précises de la consommation et des biens des ménages à partir d'images satellites diurnes. Par ailleurs, sur la base d'un modèle d'apprentissage profond,

Sirignano et al. (2016) ont étudié le risque des prêts hypothécaires créés aux États-Unis entre 1995 et 2014.

Nous adoptons donc ces méthodes afin de mieux capter l'influence des récits de demande de prêts sur la durée des campagnes. Dans la sous-section 4.3.4.5 nous présentons le principe de l'algorithme génétique, une approche à laquelle nous recourrons pour définir la structure nos deux réseaux de neurones.

Toutes les prédictions mentionnées ici ont été réalisées dans Python avec la fonction Pycaret.

4.3.4.1. La régression pénalisée

Nous utilisons la régression linéaire et logistique avec la régularisation ElasticNet pour estimer le résultat du traitement et le score de propension, et cela en incluant les 100 vecteurs obtenus à partir des récits de demande de prêts dans un premier temps et dans un second temps en les excluant afin d'appréhender l'apport de cette variable dans l'analyse. Étant donné que nous avions déjà présenté plus haut la régression linéaire avec la régularisation ElasticNet, nous ne présentons ici que la régression logistique.

Notre objectif est d'estimer la probabilité que Z prenne la valeur 1 étant donné les caractéristiques X, soit $p = \mathrm{P}(Z = 1/X = x)$ et par la suite déduire la probabilité que Z prenne la valeur 0 conditionnellement à X.

La transformation logistique, nous permet d'écrire cette probabilité sous la forme suivante :

$$\log \frac{p}{1-p} = X\beta$$

Plus encore,

$$p = \frac{e^{X\beta}}{1 + e^{X\beta}} = \frac{1}{1 + e^{-X\beta}}$$

Pour chaque projet i, la fonction de log-vraisemblance se définit comme :

$$\mathcal{L} = \sum_{i=1}^{n} [z_i \ln(p_i) + (1 - z_i) \ln(1 - p_i)]$$

Dans le cas de la régression logistique régularisée utilisant la pénalité ElasticNet, les coefficients β s'obtiennent en minimisant l'expression suivante :

$$\hat{\beta} = arg\min_{\beta}\left\{\sum_{i=1}^{n}[z_i \ln(p_i) + (1 - z_i)\ln(1 - p_i)] + \lambda_2|\beta|^2 + \lambda_1|\beta|_1\right\}$$

Pour déterminer λ_1 et λ_2 nous procédons de la même manière que dans le cas de la régression linéaire pénalisée.

4.3.4.2. La méthode des forêts aléatoires

Proposée par L. Breiman (2001), cette méthode est l'une des techniques d'apprentissage automatique les plus populaires pour les problèmes de classification et de régression à cause de sa classification discriminante efficace (Ho, 1998). Elle est plus performante que les classificateurs tels que l'analyse discriminante ou le CART car elle est plus robuste au bruit et permet de réduire le surajustement (Breiman, 2001). Le principe de la méthode consiste, en fait, à combiner plusieurs prédicteurs en arbres de classification. Chaque arbre est obtenu en échantillonnant de manière aléatoire un sous-ensemble de l'ensemble des données. Ainsi, seul un sous-ensemble de prédicteurs est utilisé pour la construction de chaque arbre, ce qui permet de réduire la corrélation entre les arbres et améliorer la qualité de l'information. Ce procédé permet également de combiner les prédicteurs des arbres qui sous-ajustent afin d'obtenir des arbres de décision plus adaptés aux données.

Le résultat de classification pour chaque observation est obtenu dans le cas de la régression en calculant simplement la moyenne des prédictions de chaque arbre. Dans le cas de la classification, c'est la règle de majorité qui est appliquée. Par exemple si la majorité des arbres de la forêt aléatoire prédit que le projet i est un projet de groupe alors cette observation sera classée comme 1 à l'issu de la classification et la probabilité associée sera estimée.

4.3.4.3. Le perceptron multicouche

En tant que variante du modèle original de Perceptron proposé par Rosenblatt en 1958, le perceptron multicouche (PMC) est l'architecture de réseau neuronal la plus simple et la plus courante. En termes de structure, il s'agit d'une combinaison de plusieurs unités regroupées en couches où chaque unité de chaque couche est connectée à chaque unité de chacune des couches suivantes. Un tel réseau comporte au moins trois couches : une couche d'entrée, une ou plus plusieurs couches intermédiaires et une couche de sortie. Le nombre d'unités de la couche d'entrée dépend du nombre des prédicteurs tandis que celui de la

couche de sortie dépend de la variable à prédire. Les couches intermédiaires sont, quant à elles, des couches cachées car les unités de ces couches sont générées lors de l'apprentissage et ne sont donc pas connues d'avance. Le nombre optimal de couches cachées et d'unités associées qu'il convient d'utiliser dépend, entre autres, du nombre d'unités dans la couche d'entrée et de sortie, de la complexité de la relation à apprendre ou de la quantité de bruit qu'il y a dans les données.

La définition de l'architecture de ce modèle n'est donc pas évidente. Cependant, elle est capitale car un modèle trop simple sera incapable de trouver la solution au problème à résoudre et ainsi être sous-ajusté aux données tandis qu'un modèle avec un réglage excessif peut se trouver être surajusté aux données (Ludermir,2005). En partant de l'architecture du réseau utilisé par Pham et Shen (2016) et en recourant à un algorithme génétique, nous sommes parvenus à définir les différents paramètres nécessaires au fonctionnement de notre modèle. Comme Pham et Shen (2016), nous choisissons de séparer les récits de demande de prêts des autres variables. Le réseau reçoit donc deux entrées séparées, une de taille 17 et une autre de taille 100. Que soit pour l'estimation des résultats de traitement ou du score de propension, nous utilisons la fonction d'activation linéaire rectifiée au niveau des couches cachées. L'usage de la fonction d'activation linéaire rectifiée dans les couches cachées est recommandé dans la littérature car elle converge rapidement et, contrairement aux fonctions telles que la fonction d'activation sigmoïde pour laquelle le gradient prend des valeurs très proches de zéro, elle permet d'équilibrer la magnitude du gradient (Le, 2015).

Dans la couche de sortie pour l'estimation des résultats de traitement, nous utilisons un seul neurone et la fonction d'activation linéaire rectifiée car la variable de réponse est numérique tandis que dans celle pour l'estimation du score de propension, nous utilisons deux neurones et la fonction d'activation Softmax. La fonction Softmax est adaptée aux problèmes de classification et fournit la probabilité qu'une observation appartienne à chaque classe considérée. Dans notre cas, nous obtenons donc pour chaque projet deux probabilités, une probabilité pour que le projet soit un projet de groupe et une autre pour qu'il ne le soit pas.

Le modèle pour l'estimation du score de propension comporte deux couches cachées, soit un total de 6 couches cachées, de plus que le modèle pour l'estimation des résultats de traitement.

Pour éviter que le modèle ne soit surajusté, nous procédons à un décrochage. Il s'agit d'une technique efficace et généralement utilisée pour contrer le surajustement dans les réseaux neuronaux (Srivastava et al., 2014). Cette technique consiste à supprimer de manière aléatoire une fraction des neurones ainsi que leurs connexions du modèle au cours de l'apprentissage.

La fonction de perte dans le cas de l'estimation du score de propension est l'entropie croisée multi-classe et dans celui de l'estimation des résultats de traitement est l'erreur quadratique moyenne. Pour ce qui est de l'optimiseur, nous optons pour Adam (adaptative momentum estimation) après un essai infructueux avec l'optimiseur de la descente de gradient stochastique.

En résumé, le modèle pour l'estimation des résultats du traitement se décrit comme suit :

Désignons par h_j^* les activations des 100 entrées reçues par la première couche d'entrée et h_j^{**} celles des 17 entrées reçues par la seconde couche d'entrée, Y l'activation des neurones de sortie et ${h_i}^{(s)}$ les neurones de la $s^{ème}$ couche cachée. Le réseau étant entièrement connecté, chaque neurone reçoit des connexions de tous les neurones de la couche précédente. Il y a donc pour chaque neurone un biais b_i et pour chaque paire de neurones dans deux couches consécutives un poids w_{ij}. Par conséquent, les calculs dans le réseau pour estimer le score de propension peuvent être écrits comme suit :

$$ {h_i}^{(1)} = f\left(\sum_j {w_{ij}}^{(1)}\, h_j^* + {b_i}^{(1)}\right) $$

$$ {h_i}^{(2)} = f\left(\sum_j {w_{ij}}^{(2)}\, {h_j}^{(1)} + {b_i}^{(2)}\right) $$

$$ {h_i}^{(3)} = f\left(\sum_j {w_{ij}}^{(3)}\, h_j^{**} + {b_i}^{(3)}\right) $$

$$ {h_i}^{(concate)} = concatener\left[{h_i}^{(2)}, {h_i}^{(3)}\right] $$

$$ {h_i}^{(4)} = f\left(\sum_j {w_{ij}}^{(4)}\, {h_j}^{(concate)} + {b_i}^{(4)}\right) $$

$$ {h_i}^{(5)} = f\left(\sum_j {w_{ij}}^{(5)}\, {h_j}^{(4)} + {b_i}^{(5)}\right) $$

$$ {h_i}^{(6)} = f\left(\sum_j {w_{ij}}^{(6)}\, {h_j}^{(5)} + {b_i}^{(6)}\right) $$

$$y_i = f\left(\sum_j {w_{ij}}^{(7)}\, {h_j}^{(6)} + {b_i}^{(7)}\right)$$

Où $f(x) = \max(0, x)$ est la fonction d'activation rectifiée.

Pour mesurer la performance de la régression, nous regardons la racine carrée de l'erreur quadratique moyenne et pour la classification nous regardons la précision et le score F1. Plus l'erreur quadratique moyenne est faible et meilleure est la prédiction des résultats de traitement. En revanche, plus la précision et le score F1 sont élevés et meilleure est la prédiction du score de propension.

4.3.4.4. Le réseau de neurones récurrent à longue mémoire à court terme

Notre objectif ici est de pouvoir capter les relations séquentielles qui existent entre les mots des récits de demande de prêts et les introduire dans l'analyse. Pour cela, nous utilisons directement les vecteurs issus des récits de demande de prêts, ce qui peut causer un problème de gradient de fuite si nous utilisons le modèle PMC. Ce problème décrit la situation dans laquelle les réseaux non récurrents (réseaux feedforward) avec de nombreuses couches ne parviennent pas à apprendre d'un jeu de données ou convergent prématurément vers une solution sous-optimale. Le réseau de neurones récurrent à longue mémoire à court terme a été conçu afin de répondre à ce problème. Dans ce type de réseau, les neurones des couches cachées des réseaux de neurones récurrents ordinaires sont remplacés par des blocs de mémoires composés d'une ou plusieurs cellules de mémoire et trois neurones spéciaux. Ces neurones spéciaux sont en fait des portes d'entrée, de sortie et d'oubli du réseau qui permettent aux cellules de mémoire de stocker et d'accéder à des informations sur de longues périodes, ce qui permet de gérer les dépendances à long terme et de résoudre le problème de gradient de fuite (Manashty et Thomson,2017). Le modèle MLCT est, par conséquent, un choix adapté dans notre cas pour avoir une prédiction plus précise et plus robuste.

Nous maintenons les mêmes architectures que dans le cas du PMC pour estimer le résultat du traitement et le score de propension. La seule différence se situe au niveau des données issues des récits de demande de prêts. En effet, dans ce modèle nous ne faisons pas d'agrégation mais nous utilisons les 100 covariables de chaque mot.

4.3.4.5. L’algorithme génétique

Dans l’apprentissage automatique, deux des plus grands et presque omniprésents problèmes sont : le sous-ajustement et le surajustement (Kolluri et al., 2020; Bashir et al., 2020; Haider, 2014). Lorsque le modèle n’est pas capable d’appréhender efficacement la variabilité qui existe dans les données alors il est sous-ajusté. Un tel modèle affichera toujours une erreur élevée aussi bien pour les données d’entrainement que pour les données de test. Quant au surajustement, il s’agit d’un phénomène opposé à celui du sous-ajustement. En d’autres mots, le modèle sur-ajusté est un modèle trop puissant au point où il apprend un peu trop bien des données d’entraînement et devient incapable de s’adapter à de nouvelles données. L’erreur sur les données d’entrainement pour un tel modèle est faible alors qu’elle est importante sur les données de test. Pour éviter ces deux problèmes, plusieurs méthodes ont été proposées. Dans le cas du sous-ajustement, il s’agit par exemple, d’augmenter la complexité du modèle ou de réduire le bruit dans les données (Castrejon et al., 2019; Reilly, 2021; Zhang et al., 2019). Pour le surajustement, nous retrouvons des méthodes telles que la pénalisation (Pereyra et al., 2017; Kolluri et al., 2020), le décrochage (Hinton et al., 2012; Srivastava et al., 2014; Zhang et al., 2019) ou l’augmentation du volume des données (Perez et Wang, 2017; Zhang et al., 2019). Khan et Allamy (2014) souligne la technique d’arrêt précoce comme une approche applicable pour réduire le sous et le surajustement. Toutes ces techniques visent à optimiser les modèles d’apprentissage automatique et conduire in fine à l’identification de l’architecture optimale.

Dans le cas particulier des réseaux de neurones, de récentes études ont montré l’efficacité des algorithmes d’optimisation comme l'algorithme d'optimisation par essaims de particules (Chen et Yu 2005) et les algorithmes évolutionnaires (Eysa et Saeed 2005 ; Harpham et al., 2004 ; Venkatesan et al., 2009 ; Yao et Islam 2008). Parmi ces derniers, on retrouve l’algorithme génétique qui s’est révélé être un outil efficace d'optimisation car il incorpore une étape de croisement en plus celle de la mutation (Hassan et al., 2016; Lins et Ludermir, 2013; Harsh et Bhatia,2012). Il s’agit d’un algorithme de recherche stochastique heuristique qui imite les principes de l’évolution génétique et de la sélection naturelle. Nous choisissons donc d’utiliser cet algorithme pour déterminer le nombre optimal de neurones dans chaque couche cachée et le nombre optimal de mots qu’il convient de retenir pour chaque récit dans le réseau de neurones récurrent à longue mémoire à court terme.

Comme l'indique la figure 1, l'algorithme génétique se déroule principalement suivant 6 étapes.

Le codage des individus :

Notre objectif est de déterminer le nombre optimal de neurones à l'intérieur de chaque couche cachée et le nombre de mots par récit pour le modèle MLCT profond. Chaque individu, dans le modèle PMC, est représenté par 6 chromosomes pour l'estimation des résultats de traitement et 4 chromosomes pour l'estimation du score de propension. Dans le modèle MLCT profond, chaque individu est représenté par 8 chromosomes pour l'estimation des résultats de traitement et 6 chromosomes pour l'estimation du score de propension. Chaque chromosome est composé de gène qui prend la valeur 1 ou 0. Par exemple, pour les neurones, une valeur 1 est attribuée si le neurone en question est activé et 0 sinon.

La population initiale :

L'algorithme débute avec une population initiale de solutions (individus) aléatoirement générées. La taille cette population initiale est un paramètre important qu'il faut choisir adéquatement car une taille trop élevée peut ralentir considérablement l'algorithme tandis qu'une taille trop faible pourrait ne pas être suffisant pour résoudre le problème. Nous optons pour une taille de 70 individus.

L'évaluation :

A la deuxième étape de l'algorithme, une valeur numérique est calculée pour chaque individu de la population à partir d'une fonction objectif appelée fonction fitness. Dans le cas des résultats de traitement, nous minimisons l'erreur quadratique moyenne et donc l'individu ayant la plus faible erreur est la meilleure solution. Pour le score de propension, nous maximisons la précision et la meilleure solution a donc la précision la plus élevée.

La sélection des parents :

La troisième étape correspond à la sélection des individus de la population initiale qui serviront de parents et s'accoupleront pour donner la prochaine génération. La méthode utilisée pour la sélection est la roulette. Elle consiste à sélectionner les individus proportionnellement à leur valeur pour la fonction fitness. Ainsi, la probabilité qu'un

individu soit sélectionné est égale au rapport entre sa valeur pour la fonction fitness et la somme des valeurs de fitness de tous les individus de la population.

Le croisement :

A cette étape, les individus sélectionnés à l'étape précédente sont utilisés en tant que parents pour créer des enfants. Nous procédons, dans le cadre de ce travail, à un croisement basé sur l'ordre avec une probabilité de 0,4. L'idée consiste à sélectionner une sous-séquence aléatoire chez le parent 1 et la copier à l'enfant 1 mais exactement au même endroit. Ensuite, il faut combler les emplacements de gènes encore vides pour l'enfant 1 par les gènes du parent 2 n'ayant pas déjà été fournis par le parent 1, tout en respectant leur ordre d'apparition chez le parent 2. Pour l'enfant 2, on procède de la même façon mais en considérant le parent précédemment parent 1 comment parent 2 et vice-versa.

La mutation :

L'objectif est d'introduire une certaine diversité dans les solutions afin d'éviter les optimums locaux. Elle consiste à modifier une ou plusieurs gènes (s) des individus choisis au hasard. Dans notre cas, nous opérons le mélange avec une probabilité de 0,6. La probabilité qu'un individu soit muté est de 0,1.

Nous utilisons le package DEAP de Python pour l'implémentation. Les résultats sont présentés à l'annexe F.

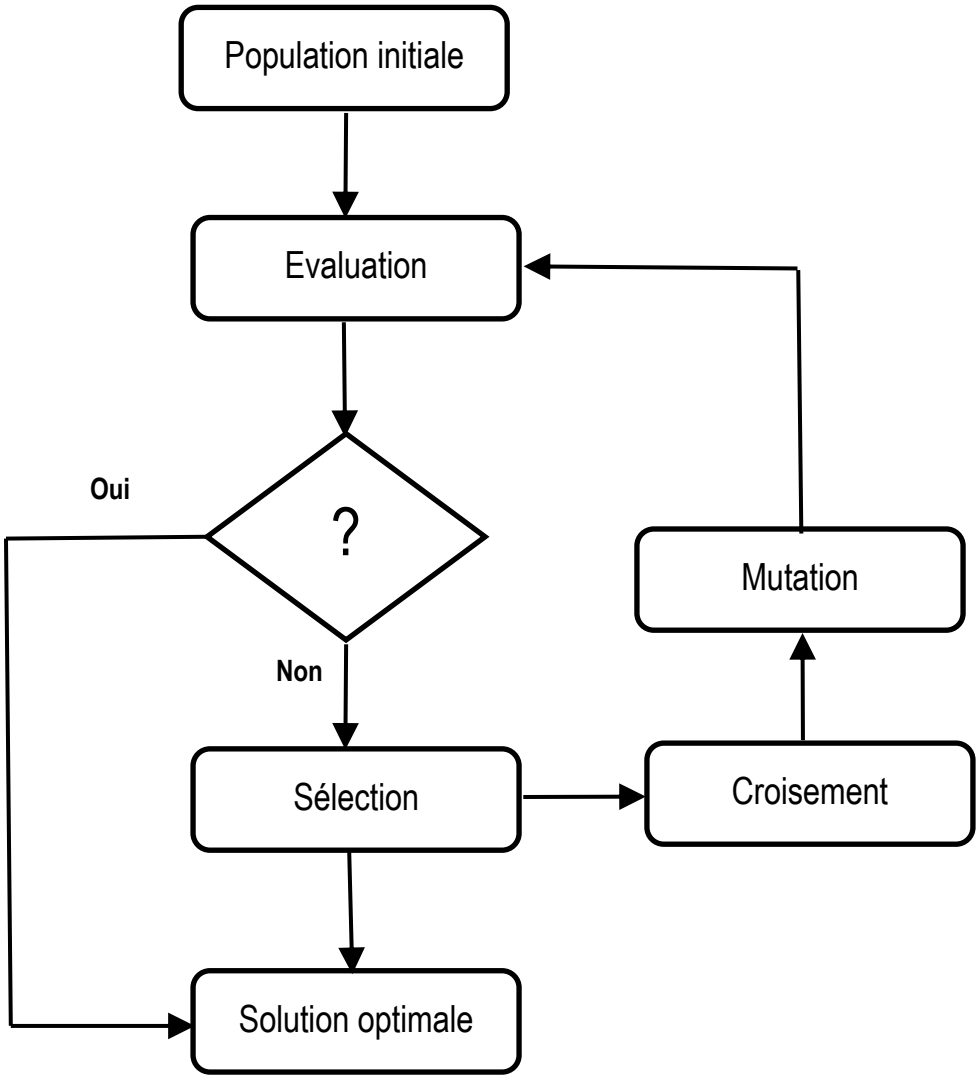

Figure 1 : Étapes de l'algorithme génétique

5. Description des données et mesures des variables

Notre terrain d'étude est la plateforme Kiva, la plus importante en matière de financement participatif à but non lucratif. Les détails sur son fonctionnement ont été présentés à la section 2 du chapitre 1 qui porte sur le financement participatif basé sur la dette (section 1.2). Dans ce chapitre, nous présentons les données de cette plateforme et les traitements que nous y faisons avant de les utiliser (section 5.1). Nous y décrivons également les variables que nous utilisons pour atteindre nos trois objectifs (section 5.2).

5.1. Données d'étude

Les données proviennent du site de Kiva, téléchargeables à partir de l'API (http://build.Kiva.org/), et couvrent la période d'avril 2006 à janvier 2019. Durant cet intervalle de temps, 1 672 913 demandes de prêts ont été publiées sur la plateforme. Pour faire correspondre chaque demande aux prêteurs qui y ont répondu, nous avons fusionné cet échantillon avec les données disponibles sur les prêteurs, ce qui a conduit à la perte de 49 524 observations soit 2,96% des demandes de prêts. Le nouvel échantillon comporte désormais 1 623 389 demandes de prêts dont 19 404 pour le mois de janvier 2019. Dans notre étude, les multi-prêteurs sont identifiés à partir des données de janvier 2019 seulement. Ce choix repose sur la difficulté à laquelle nous sommes confrontés lorsque nous essayons d'identifier les multi-prêteurs dans un échantillon de plus grande taille.

Parmi les prêteurs ayant contribué aux 19 404 demandes de janvier 2019, nous avons identifié ceux qui sont des multi-prêteurs, c'est-à-dire les prêteurs qui ont soutenu au moins 10 prêts durant ce mois. Ils sont au total 193 prêteurs. Ensuite, nous avons récupéré les demandes auxquelles ces 193 prêteurs ont contribué, soit un total de 996 989 demandes de prêts qui va d'octobre 2006 à janvier 2019. Les multi-prêteurs ont donc contribué à 61,41% des 1 623 389 demandes de prêts.

Nous avons donc, à cette étape du traitement des données, trois échantillons de données : un premier échantillon (Éch1) composé de 1 620 244 demandes de prêts s'étalant sur la période d'avril 2006 à janvier 2019, un deuxième échantillon (Éch2) composé de 996 989 demandes de prêts s'étalant sur la période d'octobre 2006 à janvier 2019 et un troisième échantillon (Éch3) composé de 626 400 demandes de prêts s'étalant sur la période d'avril

2006 à janvier 2019. Ce troisième échantillon regroupe les demandes de prêts pour lesquelles les multi-prêteurs n'ont pas contribué.

Pour étudier le niveau de diversification des portefeuilles, nous utilisons le deuxième échantillon (Éch2), c'est-à-dire les données de prêts des multi-prêteurs. De ces données, nous excluons les prêts pour lesquels nous ne disposons pas d'informations sur les prêteurs. Les données de 38 prêteurs ont donc été éliminées, ce qui nous ramène à un échantillon (Éch2a) de 155 portefeuilles de prêts, regroupant 1 791 898 prêts ou 891 171 projets.

Pour l'étude du lien entre le langage intrinsèque et extrinsèque et la durée des campagnes de financement, nous utilisons les demandes de prêts qui ont été totalement financées dans chacun des 3 échantillons (Éch1, Éch2 et Éch3). Ce qui nous ramène aux échantillons de taille respective 1 541 015 prêts (soit 95,11% des 1 620 244 demandes de prêts), 928 504 prêts (soit 93,13% des 996 989 demandes de prêts) et 612 511 prêts (soit 97,78% des 626 400 demandes de prêts). Puisque nous nous intéressons aux projets d'entrepreneuriat, nous excluons donc les prêts qui ne sont pas demandés dans le but d'entreprendre telles que les prêts destinés à un usage personnel, ceux demandés pour payer les frais de scolarités, les frais d'hospitalisation ou de rénovation de maison. Les prêts qui ont été demandés directement sans l'intermédiaire de partenaires de terrain et ceux pour lesquels nous ne disposons pas d'information sur la cote de risque du partenaire de terrain sont aussi exclus des échantillons.

Au bout du rouleau, les échantillons finaux (Éch1b, Éch2b et Éch3b) sont composés respectivement de 1 294 325 prêts, 787 189 prêts et 507 136 prêts. Les multi-prêteurs ont contribué à plus de la moitié de l'ensemble des prêts considérés soit 60,82%.

Dans l'analyse sur les groupes de prêts, nous utilisons l'échantillon Éch2. De cet échantillon, nous éliminons les prêts pour lesquelles nous ne disposons pas d'informations sur le genre de l'emprunteur ou les récits de demande de prêts, ce qui donne un échantillon final Éch2c composé de 898 772 prêts pour lesquels les multi-prêteurs ont apporté leur contribution.

Il faut noter que nous divisons aléatoirement cet échantillon en un échantillon d'apprentissage représentant 70% et un échantillon de test représentant 30%. Nous utilisons l'échantillon d'apprentissage pour les estimations et, l'échantillon de test pour les prédictions et le calcul des métriques pour comparer les méthodes. Durant le processus

d’apprentissage, 5% de l’échantillon servira à choisir les meilleurs paramètres pour chaque modèle.

5.2. Définition et mesure des variables

Dans cette section, nous définissons et décrivons comment nos variables sont mesurées. Nous débutons par les variables dépendantes (5.2.1) et par la suite nous abordons les variables indépendantes (5.2.2). Pour finir, nous décrivons les variables de contrôle (5.2.3).

5.2.1. Variables dépendantes

Dans cette étude, nous définissons deux variables dépendantes : *Duree* et *Succes*. La première a été calculée comme le nombre de jours qui séparent la date à laquelle la demande de prêt a été postée sur la plateforme Kiva et celle à laquelle elle a été entièrement financée par les prêteurs afin de mesurer la durée des campagnes de financement. Cette mesure indique l'attractivité de chaque projet (Galak et al., 2011 ; Ly et Mason, 2012b ; Allison et al., 2013). Dans les recherches sur l'entrepreneurial et le microcrédit, cette variable est utilisée comme indicateur de la préférence des prêteurs et donc de la performance entrepreneuriale (Chatterji, 2009). Pour les modèles de prêt « tout-ou-rien », elle a une importance capitale aussi bien pour les entrepreneurs que pour les investisseurs car lorsque le prêt n'est pas entièrement financé, les investisseurs n'ont pas l'opportunité d'investir et les emprunteurs n'obtiennent pas de financement.

La seconde variable dépendante mesure l'issu des campagnes de collectes de fonds. Il s'agit d'une variable binaire qui prend la valeur 1 si le montant collecté au cours de la campagne de financement d'un projet est au moins égal au montant recherché par l'emprunteur pour la réalisation de son projet et 0 sinon.

5.2.2. Variables indépendantes

Les variables indépendantes sont au nombre de quatre principalement : *Classe, Intrinseque*, *Extrinseque* et *Groupe*. Nous utilisons *Classe* dans l'étude du lien entre le niveau de diversification et la durée des campagnes de financement des projets. *Intrinseque* et *Extrinseque* sont des variables qui mesurent respectivement le niveau de langage intrinsèque et de langage extrinsèque dans les récits de demande de prêts. Il s'agit, en fait, de scores dont les valeurs élevées indiquent un niveau élevé du langage considéré et des valeurs faibles un niveau faible (Annexe A).

Nous mesurons *Intrinseque* et *Extrinseque* comme suit :

$$Intrinseque = Diversite^{*} + Interet\ humain^{*}$$

$$Extrinseque\ = Risque^{*} + Profit^{*}$$

Où $Risque^{*}$, $Profit^{*}$, $Diversite^{*}$ et $Interet\ humain^{*}$désignent respectivement les valeurs standardisées (centrées et réduites) des variables *Risque*, *Profit*, *Diversité* et *Interet_humain*. Les variables *Profit* et *Risque* mesurent le niveau du langage de profit et le niveau du langage de risque présents dans les récits et donnent un aperçu du montant de profit que le porteur de projet anticipe pour son projet ainsi que le risque qui est associé. Elles ont été mesurées grâce aux mots de profit et de risque précédemment validés et utilisés pour les recherches sur l'orientation du marché (Zachary et al., 2011).

Les variables *Interet_humain* et *Diversité,* quant à elles, mesurent respectivement l'importance accordée par l'emprunteur aux personnes et à leurs activités, et l'importance accordée par l'emprunteur aux groupes et aux différences individuelles. Elles sont mesurées à partir des dictionnaires de mots d'intérêt humain et de mots de diversité développés et validés par Hart (1984, 2001, 2010).

Pour mesurer ces quatre variables intermédiaires, nous avons utilisé le programme d'analyse de contenu DICTION 7.0 (Hart et Jarvis, 1997). Il s'agit d'un programme assez avancé qui a fait ses preuves dans des études sur le leadership (Bligh et al., 2004a et 2004b), la psychologie et des études en communication (Yuthas et al., 2002). Chacune des variables sont calculées en divisant le nombre de mots différents du dictionnaire concerné par le nombre total composant chaque récit. Étant donné que la taille des récits varie, nous avons opté pour l'option *Average*. Pour cela, chaque récit de demande de prêt a été subdivisé en des unités de 500 mots. Par la suite, seront calculés un score pour chacune des

unités obtenues et une moyenne arithmétique de ces scores. Par exemple, pour un récit avec 1 500 mots nous aurons 3 unités de 500 mots et donc 3 scores puis une moyenne arithmétique. Les mots de chacun des quatre dictionnaires sont présentés dans l'annexe A.
La variable indépendante *Groupe* est une variable binaire qui prend la valeur 1 si l'emprunteur est un groupe et 0 sinon. Nous l'utilisons dans l'analyse sur les prêts de groupe.

5.2.3. Variables de contrôle

Pour servir de variables de contrôle, nous sélectionnons sur la base des études antérieures des variables qui caractérisent l'emprunteur, son projet et sa demande de prêt car elles constituent des éléments susceptibles d'influencer la performance entrepreneuriale (Makino et al., 2004; Allison et al., 2015). Les emprunteurs viennent de 88 pays différents, ce qui nous amène à créer pour la variable *Pays* 87 variables factices. La variable *Secteur*, comporte les 15 secteurs d'activité utilisés par Kiva pour les activités des emprunteurs. Nous utilisons le secteur de l'agriculture comme la modalité de référence et créons 14 variables factices pour les autres modalités. La plupart des demandes de prêts effectuées sur Kiva transitent par les partenaires sur le terrain et en tant qu'intermédiaires leur niveau de risque est susceptible d'influencer le choix d'investissement des prêteurs. Nous mesurons leur niveau de risque à partir de la cote de risque que Kiva associe à chaque partenaire. Cette cote indique la probabilité de remboursement des prêts octroyés par l'intermédiaire des partenaires. Selon les résultats de Priberny et Sommer (2021) une notation du risque entre 4 et 5 augmente la probabilité de financement de 0,69 %. Il existe une autre source de risque associée aux prêts qui sont effectués dans une devise autre que celle du prêteur et ce risque est dans certain cas supporté par le partenaire de terrain, le prêteur ou les deux. Selon que ce risque soit couvert par le partenaire en totalité ou en partie, cela pourrait influencer sur les décisions de prêts. Nous mesurons donc cette source de risque avec la variable à trois modalités *Couverture* qui prend la valeur 0 lorsqu'aucune perte n'est attendue, 1 si une perte en-dessous de 10% du principal est attendue et qu'elle sera supportée par le partenaire de terrain et 2 si la perte attendue est supérieure à 10% et sera supportée par le prêteur et le partenaire de terrain. Nous contrôlons aussi pour deux autres sources potentielles de risque : le nombre de mois que prendra le remboursement de l'emprunt et le montant de l'emprunt. Le montant de l'emprunt est mentionné dans la

littérature comme un facteur susceptible d'impacter sur le succès des campagnes de financement (Galak et al., 2011). Les résultats de l'étude de Mollick (2014) rejoignent ceux de Herzenstein et al. (2008) révélant que le montant de prêt demandé par les porteurs de projets agit négativement sur le succès du financement.

Allison et al. (2013) ont examiné si et dans quelle mesure l'utilisation d'un langage politique dans les récits de demande de prêts (accomplissement, blâme, préoccupation actuelle, nivellement, la ténacité et la variété) influence sur les décisions d'investissement en microcrédit. Selon les auteurs, le recours aux accomplissements, au blâme et le fait de parler des préoccupations d'actualité dans les récits permettent aux prêteurs d'évaluer à quel point le besoin de l'emprunteur est pressant. En nivelant et en rendant moins concret son récit, l'emprunteur peut permettre aux prêteurs de s'identifier à lui et ainsi financer son projet. Enfin, la clarté d'un récit de prêt, mesuré au-travers de la ténacité et la répétition de certains points essentiels, permet aux prêteurs de cerner l'objectif de la demande de prêt. Les auteurs parviennent à la conclusion selon laquelle, les récits qui sont encrés sur le blâme et les préoccupations d'actualité se financent plus vite tandis que les récits qui sont encrés sur les accomplissements, la ténacité et la variété se financent plus lentement. Nous ajoutons donc ces sept variables pour servir de contrôle et les mesurons à partir des dictionnaires de mots disponibles dans DICTON 7.0.

Lorsque l'emprunteur est une femme alors la variable *Genre* prend 1 et 0 s'il s'agit d'une femme. Pour les groupes, nous attribuons la valeur 1 si le nombre de femmes est supérieur ou égal au nombre d'hommes et 0 si le nombre de femmes est inférieur au nombre d'hommes. Dans le cas où le risque lié au prêt est supporté par le prêteur alors la variable *Risque_defaut* prend la valeur 1 et 0 s'il est supporté par le partenaire de terrain.

Nous utilisons également les récits de demande de prêts comme variable de contrôle dans l'analyse des prêts de groupe.

Un exemple de récit pour un prêt individuel se présente comme :

Florence is married with four children. She has been in the group for 4 years. Currently she is the group treasurer. She is a treasurer of Revelations women group and another self help group. They buy household utensils for members and distribute amongst themselves. Florence keeps dairy cows and sells milk at Landless and Thika town. She has also a retail

shop in Gatundu-Makongeni area. Due to high demand of milk, she would like to increase supply by buying more milk from vendors as well as buying another dairy cow.

Pour un prêt de groupe, un exemple est :

This group of farmers lives in Bluangan, Perean Kangin village, Baturiti, Tabanan. This group led by I Wayan S. comprises I Gede Made Sukardi, I Nyoman Subawa, I Ketut Sucita dan I Made Juniarta. As cabbage farmers, they require at least 3 months, from preparing the land for planting cabbage seeds to harvesting. They prepare the land through mutual cooperation. Before the field is ready for preparation, they are aerated. This will make the field easier for the process. When the fields and seed are ready they are planted. Thereafter commences the care. They need fertilizer for nutrition and pesticide for the insects. The amount of water also needs to be controlled, not too much or too little. These activities are carried out till almost harvest. Their effort will generate a profit of Rp 3.000.000,00 (about USD330) per person. This amount is attained from 3 months of farming. Apart from cabbage, they also grow padi and tomato. Their loan will be used to purchase seeds, manure, and pay for care and working expenses.

Pour transformer ces données textuelles en des données numériques, nous utilisons l'intégration des mots. L'idée dans cette approche est de représenter chaque mot employé dans les récits comme un vecteur de nombres réels de faible dimension en s'appuyant sur le sens et le contexte du mot. Les méthodes les plus récentes et les plus efficaces pour accomplir cette tâche sont Word2Vec et GloVe. Le choix entre ces deux méthodes dépend essentiellement des données. Selon Pennington et al. (2014), Word2Vec est moins efficace que GloVe car il n'utilise pas toutes les informations du corpus d'entraînement. Cependant Levy et al. (2015) aboutissent à l'opposé. Pour eux, cette contradiction dans les résultats est due, entres autres, au fait que Pennington et al. (2014) n'ont utilisé que les analogies de Google pour effectuer leur comparaison. Pham et Shen (2015), quant à eux, ont utilisé GloVe pour effectuer leur analyse. Dans cette étude, nous utilisons l'incorporation de pré-entraîné de GloVe de dimension 100. Ainsi, pour chaque mot nous un vecteur composé de 100 nombres réels correspondant chacun a une dimension.

Nous utilisons directement ces représentations ainsi obtenues, pour chaque mot, dans le réseau de neurones récurrent à longue mémoire à court terme, ce qui nous donne pour chaque mot un vecteur de 100 nombres réels. Dans le perceptron multicouche, nous

calculons la moyenne des représentations des mots contenus dans les récits de demande de prêts qui sont fournis pour chaque demande de prêts, ce qui nous ramène à un vecteur de 100 nombres réels pour chaque prêt et non pour chaque mot.

6. Analyse descriptive

Le tableau 1 présente la liste des pays des multi-prêteurs ainsi que le nombre de contributions effectuées par pays. Les multi-prêteurs viennent de 23 pays différents et seulement 7 de ces pays regroupent plus de deux prêteurs. Même si les multi-prêteurs venant des Etats-Unis d'Amérique sont les plus nombreux et ont le nombre le plus élevé de contributions, ils n'ont cependant pas le nombre plus élevé de contributions par tête. La Norvège est le pays avec le nombre de contribution par tête le plus élevé. Un autre pays qui attire l'attention est le Luxembourg qui a le plus petit nombre de contributions par tête.

Tableau 1 : Nombre de multi-prêteurs par pays, nombre de contributions par pays et per capita[7]

Pays	Nombre de multi-prêteurs	Nombre de contributions	Nombre de contributions per capita
Etats-Unis d'Amérique	90 (58,06%)	1071413 (59,79%)	11905
Norvège	10 (6,45%)	171448 (9,57%)	17145
Pays-Bas	10 (6,45%)	116599 (6,51%)	11660
Royaume-Uni	7 (4,52%)	115563 (6,45%)	16509
Allemagne	7 (4,52%)	83561 (4,66%)	11937
Canada	6 (3,87%)	42484 (2,37%)	7081
Autriche	4 (2,58%)	42088 (2,35%)	10522
France	3 (1,94%)	27047 (1,51%)	9016
Suisse	2 (1,29%)	26416 (1,47%)	13208
Suède	2 (1,29%)	25979 (1,45%)	12990
Dominique	2 (1,29%)	14772 (0,82%)	7386
Iran	1 (0,65%)	11557 (0,64%)	11557
Émirats arabes unis	1 (0,65%)	8929 (0,50%)	8929
Danemark	1 (0,65%)	7053 (0,39%)	7053
République tchèque	1 (0,65%)	6590 (0,37%)	6590
Irlande	1 (0,65%)	5533 (0,31%)	5533
Japon	1 (0,65%)	4413 (0,25%)	4413
Porto Rico	1 (0,65%)	3649 (0,20%)	3649
Arabie Saoudite	1 (0,65%)	2080 (0,12%)	2080
Thaïlande	1 (0,65%)	1836 (0,10%)	1836
Taïwan	1 (0,65%)	1251 (0,07%)	1251
Finlande	1 (0,65%)	1227 (0,07%)	1227
Luxembourg	1 (0,65%)	410 (0,02%)	410
Total	**155 (100%)**	**1 791 898 (100%)**	**11 561**

[7] Per capita fait référence à « Par nombre de multi-prêteurs de ce pays dans l'échantillon »

En regardant la liste des pays des emprunteurs (Annexe C), on remarque que presque tous les continents sont représentés (Afrique, Europe, Amérique, Océanie, Asie). Les demandes de prêts effectuées par les emprunteurs philippiens représentent la part la plus élevée dans notre échantillon, soit 15,43%. Les deux autres pays avec des parts au-delà de 5% sont le Kenya et le Pérou, soit respectivement 11,2 % et 5,05 %. Malgré la courte période que couvrent les données de Allison et al. (2015) par rapport aux nôtres, les Philippines, le Kenya et le Pérou ont également les parts les plus importantes dans leur échantillon. Ils trouvent respectivement 22,05%, 14,62% et 10,79%. Pour les pays comme l'Uruguay, le Botswana, la Mauritanie ou le Bangladesh, une seule demande de prêt a été enregistrée.

Le tableau 2 fournit des statistiques sur les portefeuilles de prêts des multi-prêteurs. Ces prêteurs ont contribué en moyenne 11 561 fois avec un écart-type de 21 926. Avec un écart-type qui fait presque le double de la moyenne, nous constatons que le nombre de projets financés par prêteurs varie considérablement autour de la moyenne. En contribuant au financement d'un projet, les multi-prêteurs portent leur choix sur un pays, un secteur d'activité, sur un type de porteur de projet (individu, groupe, autre) et sur un genre (féminin, masculin et autre). En moyenne, les multi-prêteurs ont choisi 14 des 15 secteurs d'activité avec un écart-type de 1. Ils ont soutenu des projets de 55 pays différents, avec un écart-type de 17. Concernant le genre du porteur de projet, les multi-prêteurs ont accordé en moyenne 8 333 fois un minimum de 25 $ à des femmes avec un écart-type de 14 896. Aux hommes, ils ont accordé en moyenne 3 011 fois un minimum de 25 $ avec un écart-type de 7 167. En ce qui concerne le type du porteur de projets, les multi-prêteurs ont accordé en moyenne 8 254 fois un minimum de 25 $ à des emprunteurs seuls avec un écart-type de 17 026. Pour les groupes, c'est en moyenne 3 089 fois que les multi-prêteurs ont accordé un minimum de 25 $ avec un écart-type de 1 308. Sans égard au genre et au type des porteurs de projets, les multi-prêteurs ont accordé en moyenne 217 fois un minimum de 25 $ avec un écart-type 466 217.

La moitié des multi-prêteurs étudiés ont accordé plus de 5 125 fois un minimum de 25 $. Au travers de leurs contributions, ils ont investi dans les 15 secteurs d'activité et dans plus de 58 pays. Ils ont contribué plus de 3 264 fois afin de financer des projets de femmes, 977 fois pour des projets portés par des hommes, 3 505 fois pour soutenir des emprunteurs

seuls, 1 308 fois pour les groupes d'emprunteurs et enfin 76 fois pour certains porteurs de projets sans connaître leur genre ni leur type.

Tableau 2 : Sommaire statistique sur la composition des portefeuilles de prêts

		Min	Moyenne	Max	Ecart-type	Médiane
Prêts accordés		53	11 561	171 448	21 926	5 125
Secteurs d'activités		7	14	15	1	15
Pays de l'emprunteur		1	55	90	17	58
Genre de l'emprunteur	Féminin	5	8 333	108 601	14 896	3 264
	Masculin	0	3011	58434	7167	977
	Autre	0	217	4413	466	76
Type de l'emprunteur	Individu	40	8 254	14 0071	17026	3 505
	Groupe	0	3 089	33 422	4951	1308
	Autre	0	217	4 413	466	76

La figure 2 donne un aperçu de la composition des portefeuilles de prêts des multi-prêteurs avec en ordonné le nombre de portefeuilles et en abscisse le nombre de catégorie de projets. En examinant cette figure, on peut constater que la majorité des portefeuilles de prêts présentent une assez bonne diversification. En effet, pour la catégorie du genre et du type de l'emprunteur, 148 des 155 portefeuilles (soit 95,48%) comportent les 3 genres (féminin, masculin, autre) et les 3 types d'emprunteurs (groupe, individu, autre). Dans la catégorie du secteur d'activité, 139 portefeuilles (soit 89,68%) comportent au moins 12 des 15 secteurs définis sur la plateforme, 12 portefeuilles (soit 7,74%) en comportent au moins 10 et 4 portefeuilles (soit 2,58%) en comportent au moins 7. Enfin, pour le pays de l'emprunteur, 70 des 155 portefeuilles (soit 45,16%) comportent les projets venant d'au moins 51 pays différents, 32 portefeuilles (soit 20,64%) comportent des projets qui viennent d'au moins 31 pays différents et 16 (soit 10,32%) d'au moins 10 pays différents.

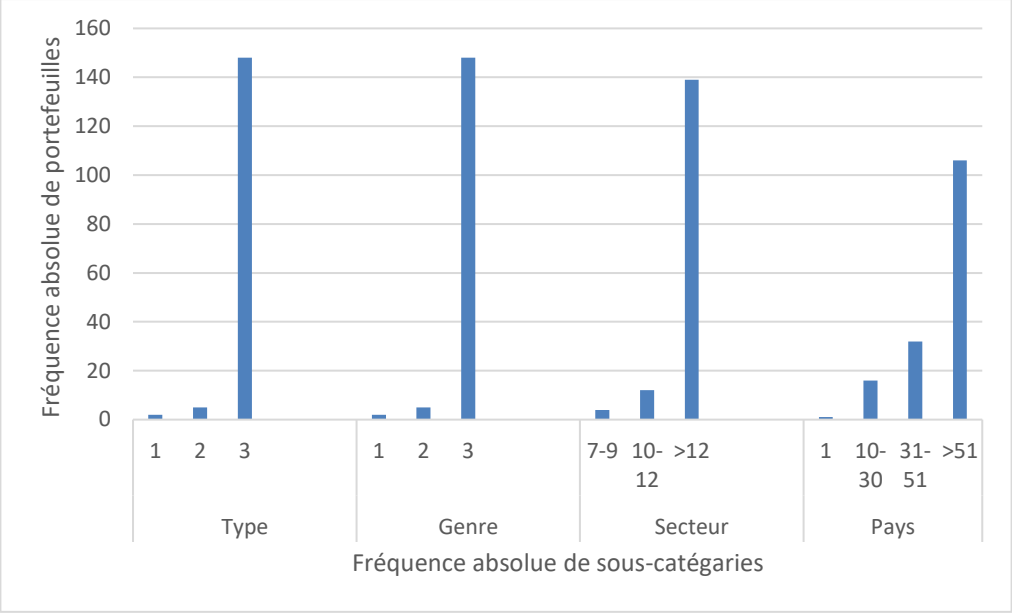

Figure 2 : Composition des portefeuilles de prêts

Le tableau 3 présente un sommaire statistique sur les variables quantitatives utilisées dans cette étude. Nous remarquons que le temps nécessaire au financement des projets est en moyenne de 12,51 jours, soit environ 12 jours, 12 heures et 14 min, avec un écart-type de 12,15 jours. La durée minimale est de 1,44 min et la plus longue de 176,6 jours. Le prêt médian s'est fait financer au bout de 7,7 jours.

Pour les indices intrinsèque et extrinsèque, les moyennes sont identiques et égales à 0,015 (elles diffèrent seulement à la cinquième décimale) avec des écart-types de 1,37 et 1,40. Le niveau de langage intrinsèque médian est de -0,12 et celui extrinsèque est de -0,30.

Pour l'échéance des prêts (Duree_pret), la moyenne est d'environ 1 an, plus précisément 12,32 mois avec un écart-type de 5,99. L'échéance la plus courte est de 2 mois et la plus longue de 146 mois ce qui représente environ 12 ans. La moitié des prêts, soit 399 368 demandes de prêts, a une échéance de plus d'un an.

Pour ce qui est des montants demandés par les emprunteurs, la moyenne se situe autour de 805 $ avec un écart-type de 987,75 $. Le montant le plus élevé est de 100 000 $ et le plus bas de 25 $. La moitié des emprunteurs ont requis un montant au-dessus de 500 $. En recourant au logarithme népérien des montants des prêts, nous obtenons une moyenne de 6,28 et un écart-type de 0,85; ces valeurs restent assez proches de celles obtenues par Allison et al. (2015) (Moyenne=6,16; écart-type=0,75).

En matière de risque, le niveau de risque des partenaires est de 2,28 en moyenne avec un écart-type de 1,70. Ce niveau était de 3,3 dans l'échantillon de Allison avec un écart-type de 0,75. Étant donné qu'une cote élevée signifie que le partenaire est fiable, nous pouvons considérer que les prêts de notre échantillon sont présentés par des partenaires avec en moyenne un niveau de risque plus important que ceux dans l'échantillon exploité par Allison et al. (2015).

Tableau 3 : Sommaire de statistiques sur les variables quantitatives

Variables	Min.	1er Qu	Médian	Moy.	Écart-type	3e Qu.	Max.
Durée	0,001	3,503	7,724	12,513	12,151	19,208	176,603
Intrinsèque	-3,603	-0,913	-0,122	0,015	1,382	0,772	23,299
Extrinsèque	-1,344	-0,863	-0,298	0,015	1,396	0,543	33,555
Montant_pret	25	300	500	805	987,749	975	100 000
Duree_pret	2	8	12	12,320	5,993	14	146
Risque_IMF	0	0	2,5	2,288	1,696	4	4,500
Accomplissement	0	23,290	30,950	32,130	12,645	39,780	142,860
Blame	0	0	0	1,118	2,390	0	25,86
Preo_actuelle	0	8,4	13,34	14,260	8,256	18,990	125
Langage de ténacité	0	23,95	31,03	31,63	11,085	38,8	112,9
Nivellement	0	0	0	2,713	3,679	4,75	50,5
Caractere_concret	0	20,83	28,3	29,810	12,936	37,38	269,02
Variete	0	0,6	0,64	0,644	0,065	0,69	1

Dans le tableau 4, nous procédons à une comparaison des moyennes par variable entre les échantillons Éch2b et Éch3b en appliquant le test de Welch. Cette comparaison a pour but de faire ressortir les différences entre les projets auxquels les multi-prêteurs ont contribué (Éch2b) et ceux auxquels ils n'ont pas contribué (Éch3b). Pour la plupart des variables, nous obtenons une différence significative. Nous voyons que contrairement aux projets de l'échantillon Éch3b, ceux de l'échantillon Éch2b se financent moins vite. Lin et Boh (2019) ont fait ce même constat. Les multi-prêteurs semblent donc cibler principalement les campagnes qui ont de la difficulté à attirer des prêteurs. Le niveau du langage intrinsèque et du langage extrinsèque dans l'échantillon Éch2b est largement supérieur à celui de l'échantillon Éch3b. Cela indique que les emprunteurs que les multi-prêteurs soutiennent

présentent leur projet à la fois comme une activité rentable mais également comme un moyen d'aider les autres. Toutefois, le langage intrinsèque est plus important que le langage extrinsèque démontrant que l'aide reste primordial pour ces prêteurs.

Une autre caractéristique sur laquelle ces deux échantillons diffèrent visiblement est le montant des prêts. Les projets de l'échantillon Éch2b requièrent en moyenne près de 400 $ de plus que les projets de l'échantillon Éch3b.

Tableau 4 : Comparaison des échantillons Éch2b et Éch3b

Variables	Éch2b Moyenne (sd)	Éch3b Moyenne (sd)	Test de Welch
Durée	14,595 (12,488)	9,282 (10,844)	***
Intrinsèque	0,053 (1,366)	-0,042 (1,403)	***
Extrinsèque	0,037 (1,403)	-0,018 (1,385)	***
Montant_pret	966,134 (1158,993)	554,598 (549,565)	***
Duree_pret	12,410 (6,125)	12,174 (5,781)	***
Risque_IMF	2,481 (1,634)	1,990 (1,747)	ns
Accomplissement	32,370 (12,630)	31,749 (12,659)	ns
Blame	1,143 (2,433)	1,080 (2,322)	***
Preo_actuelle	14,333 (8,307)	14,155 (8,176)	***
Langage de ténacité	31,931 (10,938)	31,169 (11,294)	***
Nivellement	2,770 (3,765)	2,623 (3,538)	***
Caractere_concret	30,189 (12,830)	29,216 (13,076)	ns
Variete	0,642 (0,064)	0,647 (0,067)	ns

Note : *** p-value< 0.001

L'échantillon Éch2c est composé de 81,54 % des prêts demandés par des emprunteurs seuls et 18,46 % de prêts demandés par des groupes d'emprunteurs (voir Tableau 5). En comparaison à l'échantillon de Pham et Shen (2016), on note une différence d'environ 4% pour les prêts demandés par les groupes. Pour le genre, il y a 76,40% des prêts qui sont demandés par des femmes et 23,60% par des hommes. Dans 92,42% des cas le risque de défaut est supporté par les partenaires de terrains et dans 7,58% des cas par les prêteurs.

Tableau 5 : Nombre de prêts par groupe et individu, genre et risque

Modalités	Groupe	Genre	Risque
0	732 831 (81,54%)	212 099 (23,60%)	830 648 (92,42%)
1	165 941 (18,46%)	686 673 (76,40%)	68 124 (7,58%)
Total	898 772 (100%)	898 772 (100%)	898 772 (100%)

Le tableau 6 fournit des statistiques relatives à la durée des campagnes de financement des prêts, les montants qui sont demandés et au temps nécessaire au financement de 25 $ US selon qu'il s'agit d'un groupe d'emprunteurs ou d'un emprunteur seul, selon le genre de l'emprunteur et également selon que le risque de défaut est supporté par le prêteur ou le partenaire de terrain.

Les prêts demandés par les groupes mettent en moyenne 14,41 jours pour leur financement avec un écart-type de 12,22. Ce temps est quasiment le même pour les prêts demandés par les emprunteurs seuls (Moyenne=14,40 et écart-type=12,50). Cependant, les emprunteurs seuls (Moyenne=0,70 et écart-type=0,68) mettent pratiquement le double du temps que mettent les emprunteurs en groupe (Moyenne=0,35 et écart-type=0,46) pour collecter 25 $ US sur la plateforme. La durée minimale et celle maximale nécessaire pour le financement des prêts dans le cas des emprunteurs seuls sont respectivement de 0,001 jour et 167,34 jours alors qu'elles sont de 0,013 jour et 176,60 jours dans le cas des groupes. Même si la durée minimale pour 25 $ US est la même pour les deux types de prêteurs, la durée maximum pour les emprunteurs seuls est supérieure à celles des groupes d'au moins 10 jours.

Le prêt médian au niveau des groupes nécessite 10,36 jours alors qu'au niveau des emprunteurs seul le prêt médian ne nécessite que 9,57 jours. Le montant moyen que demandent les groupes (Moyenne=2017,55 $ US et écart-type=1697,18) est largement supérieur au montant moyen des prêts des emprunteurs seuls (Moyenne=703,82 $ US et écart-type=768,28). Les montants minimal et maximum sont respectivement de 50 $ US et 50 000 $ US pour les groupes mais de 25 $ US et 10 000 $ US pour les emprunteurs seuls. Pour le prêt médian, nous avons un montant de 1450 $ US pour les groupes et un montant de 500 $ US pour les emprunteurs seuls. Cependant, pour collecter 25 $ US, les emprunteurs seuls (Moyenne=0,35 jour et écart-type=0,46) mettent plus de temps que les groupes (Moyenne=0,70 jour et écart-type=0,68).

Contrairement aux prêts demandés par les hommes (Moyenne=13,06 et écart-type=11,64), les prêts demandés par les femmes (Moyenne=18,75 et écart-type=13,90) se financent plus vite. La durée minimum pour les femmes est de 0,01 jour et de 0,005 pour les hommes. Le prêt médian s'est financé en 8,53 jours pour les femmes et 19,94 jours pour les hommes. La différence entre le montant moyen des prêts demandés par les femmes (Moyenne=942,17 $ US et écart-type=1138,25) et celui des hommes (Moyenne=960,02 $ US et écart-type=1095,05) est d'environ 20 $ US. Les montants minimums sont identiques et égaux à 25 $ US tandis que les montants maximums diffèrent de 100 $ US pour les hommes. La moitié des prêts demandés par les femmes a été pour un montant au-dessus de 575 $ US et pour les hommes au-dessus de 725 dollars. Pour collecter 25 $ US sur la plateforme, les femmes mettent en moyenne 0,60 jour et les hommes 0,74 jour. Les écart-types sont de 0,64 et 0,72. Pour le prêt médian, nous avons une durée moyenne de 0,38 jour pour les femmes et 0,54 jour pour les hommes. Il y a une différence d'environ 5 jours en faveur des femmes pour le maximum de la durée nécessaire à la collecte de 25 $ US. Lorsque les prêteurs supportent le risque de défaut, la durée moyenne (Moyenne=11,61 et écart-type=12) pour le financement est moindre que quand le risque est supporté par les partenaires (Moyenne=14,63 et écart-type=12,46). Cette tendance est la même pour la collecte de 25 $ US et la différence est d'environ 0,18 jour.

Tableau 6 : Sommaire statistique de la durée et des montants en fonction du groupe, genre et risque

Variables		Groupe		Genre		Risque	
		Groupe=1	Groupe=0	Femme	Homme	Risque =0	Risque=1
Duree	Minimum	0,013	0,001	0,001	0,005	0,001	0,013
	1er Qu	4,857	4,470	4,181	6,434	4,702	2,646
	Médian	10,357	9,572	8,532	15,944	9,989	6,755
	Moyenne	14,415	14,398	13,057	18,754	14,631	11,606
	Écart-type	12,2166	12,50	11,640	13,899	12,457	12
	3e Qu.	22,569	23,833	20,430	30,216	24,131	17,458
	Maximum	176,603	167,339	176,603	167,339	176,603	78,610
Montant	Minimum	50	25	25	25	25	50
	1er Qu	750	300	325	425	325	400
	Médian	1450	500	575	725	600	675
	Moyenne	2017,553	703,824	942,166	960,020	945,055	962,523
	Écart-type	1697,182	768,28	1138,254	1095,054	1143,961	914,819
	3e Qu.	2950	925	1050	1150	1075	1100
	Maximum	50 000	100 000	100 000	50 000	100 000	7 975
Duree (Pour 25$)	Minimum	0	0	0	0,001	0	0,001
	1er Qu	0,072	0,211	0,152	0,235	0,176	0,091
	Médian	0,184	0,486	0,384	0,536	0,433	0,251
	Moyenne	0,351	0,697	0,602	0,736	0,647	0,471
	Écart-type	0,464	0,685	0,643	0,716	0,666	0,612
	3e Qu.	0,455	0,958	0,823	0,997	0,887	0,625
	Maximum	9,097	19,198	14,221	19,198	19,198	12,979

Le contenu du tableau 7 donne une idée de la manière dont les montants de prêts affectent la durée des campagnes de financement pour les projets de groupe et les projets individuels. Pour les deux types d'emprunteurs, le montant recherché se trouve être significativement et positivement corrélé à la durée des campagnes de financement. Lorsque le montant de prêt augmente de 10 000 $ US, la durée des campagnes pour les projets de groupe augmente de 19,16 jours alors qu'elle augmente de 42,12 jours pour les projets individuels. Pour un montant de 425 $ US, les projets de groupe et les projets individuels mettent le même nombre de jours pour leur financement, soit 13,415 $ US. Au-delà de ce montant, la durée des campagnes de financement des projets de groupe est inférieure à celle des projets individuels tandis qu'en dessous de ce montant c'est l'inverse. Il semble donc que les demandes de groupe avec des montants de prêts importants sont considérées par les

prêteurs comme normales contrairement aux demandes individuelles avec des montants élevés.

Tableau 7 : Résultats de la régression du montant de prêts sur la durée des campagnes

	Coefficients		R^2 (ajusté)
	Constante (Erreur standard)	Montant du prêt (Erreur standard)	
Groupe	13,16*** (0.05)	0,0006*** ($1.76x10^{-05}$)	0.015
Individu	12,14*** (0.02)	0,003*** ($1.86x10^{-05}$)	0.048

Note: *** p-value< 0.001, ** p-value< 0.01, * p-value< 0.05, +p-value< 0.1

La figure 3 présente comment les demandes de prêts se répartissent en fonction des secteurs d'activité. Les secteurs de l'agriculture (27,71%), de l'alimentation (20,94%) et du commerce de détail (19,48%) affichent les parts les plus élevées de prêts financés mais les secteurs de divertissement (0,11%) et de commerce de gros (0,11%) en affichent les moins élevées.

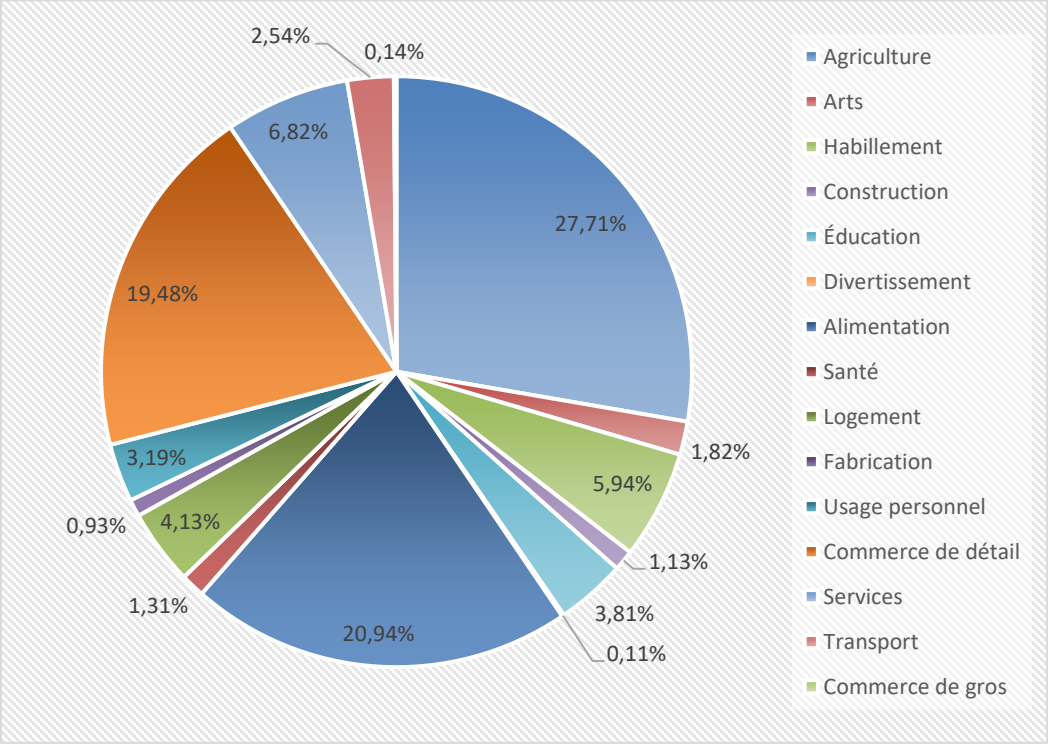

Figure 3 : Pourcentage de prêts par secteur d'activité

Dans le tableau 8, nous croisons la variable *Secteur* avec les variables *Groupe*, *Genre* et *Risque*. Il apparaît que les secteurs dominants sont l'agriculture, l'alimentation et le commerce de détail tandis que le commerce en gros et le divertissement sont les secteurs avec les plus faibles parts.

Tableau 8 : Nombre de prêts par groupe, genre et risque en fonction des secteurs d'activité (Fréquence relative en %)

	Groupe		Genre		Risque	
	Groupe=0	Groupe=1	Homme	Femme	Risque =0	Risque=1
Agriculture	200 692 (27,39%)	48 318 (29,12%)	79 217 (37,35%)	169 793 (24,73%)	235 599 (28,36%)	13 411 (19,69%)
Arts	13 058 (1,78%)	3 334 (2,01%)	1 925 (0,91%)	14 467 (2,11%)	15 461 (1,86%)	931 (1,37%)
Habillement	39 989 (5,46%)	13394 (8,07%)	5 255 (2,48%)	48 128 (7,01%)	47 435 (5,71%)	5 948 (8,73%)
Construction	8 891 (1,21%)	1 237 (0,75%)	5 702 (2,69%)	4 426 (0,64%)	9 237 (1,11%)	891 (1,31%)
Éducation	32 457 (4,43%)	1 745 (1,05%)	14 222 (6,71%)	19 980 (2,91%)	33 780 (4,07%)	422 (0,62%)
Divertissement	926 (0,13%)	85 (0,05%)	498 (0,23%)	513 (0,07%)	873 (0,11%)	138 (0,2%)
Alimentation	148 218 (20,23%)	39 975 (24,09%)	24 935 (11,76%)	163 258 (23,78%)	170 030 (20,47%)	18 163 (26,66%)
Santé	10 298 (1,41%)	1479 (0,89%)	4 119 (1,94%)	7 658 (1,12%)	11 377 (1,37%)	400 (0,59%)
Logement	34 950 (4,77%)	2 178 (1,31%)	10 262 (4,84%)	26 866 (3,91%)	35 908 (4,32%)	1 220 (1,79%)
Fabrication	7 313 (1%)	1 032 (0,62%)	3 897 (1,84%)	4 448 (0,65%)	7 678 (0,92%)	667 (0,98%)
Usage personnel	14 723 (2,01%)	13990 (8,43%)	9 349 (4,41%)	19 364 (2,82%)	28 408 (3,42%)	305 (0,45%)
Commerce de détail	146 192 (19,95%)	28 883 (17,41%)	24 288 (11,45%)	150 787 (21,96%)	156 964 (18,9%)	18 111 (26,59%)
Services	52 411 (7,15%)	8908 (5,37%)	15 774 (7,44%)	45 545 (6,63%)	56 585 (6,81%)	4 734 (6,95%)
Transport	21 664 (2,96%)	1 201 (0,72%)	12 174 (5,74%)	10 691 (1,56%)	20 223 (2,43%)	2 642 (3,88%)
Commerce de gros	1 049 (0,14%)	182 (0,11%)	482 (0,23%)	749 (0,11%)	1 090 (0,13%)	141 (0,21%)

Total	732 831 (100%)	165 941 (100%)	212 099 (100%)	686 673 (100%)	830 648 (100%)	68 124 (100%)

Le tableau 9, quant à lui, nous donne des statistiques relatives à la durée des campagnes de financement des prêts, les montants qui sont demandés et au temps nécessaire au financement de 25 $ US en fonction des secteurs d'activité.

Les durées des campagnes sont en moyenne plus élevées dans les secteurs du transport (Moyenne=17,89 et écart-type=13,71), de logement (Moyenne=16,91 et écart-type=13,67) et de l'éducation (Moyenne=9,68 et écart-type=9,15). Les projets d'arts (Moyenne=8,63 et écart-type=8,09) et de fabrication (Moyenne=8,74 et écart-type=8,06) ont les durées moyennes les plus courtes. Les montants les plus élevés sont observés pour les projets de commerce en gros (Moyenne=1437,39 $ US et écart-type=2522,05), divertissement (Moyenne=1211,15 $ US et écart-type=1277,37), l'habillement (Moyenne=1203,87 $ US et écart-type=1360,69). Pour ce qui est du temps nécessaire pour obtenir 25 $ US, le secteur du divertissement (Moyenne=0,32 et écart-type=0,34) affiche la durée la plus courte, suivie du secteur de la fabrication (Moyenne=0,33 et écart-type=0,38) et du secteur de l'éducation (Moyenne=0,35 et écart-type=0,41). Les secteurs pour lesquels cette collecte de 25 $ US est plus lente sont le secteur de l'usage personnel l'éducation (Moyenne=0,88 et écart-type=0,92), le secteur du transport l'éducation (Moyenne=0,808 et écart-type=0,73) et le secteur du logement l'éducation (Moyenne=0,73 jour et écart-type=0,69).

Tableau 9 : Moyennes et Écart-types de la durée et des montants par secteur d'activité

	Durée		Montant		Durée (pour 25$)	
	Moyenne	Ecart-type	Moyenne	Ecart-type	Moyenne	Ecart-type
Arts	8,633	8,090	1159,139	1431,552	0,386	0,490
Industrie manufacturière	8,736	8,060	1031,633	1166,179	0,331	0,384
Éducation	9,683	9,149	1063,721	1129,972	0,346	0,411
Divertissement	10,492	9,229	1211,152	1277,370	0,318	0,345
Usage personnel	11,159	11,481	532,278	680,704	0,880	0,918
Commerce de gros	12,965	11,179	1437,388	2522,047	0,369	0,419
Santé	12,970	11,102	1117,740	1251,960	0,458	0,477
Alimentation	13,606	12,132	1002,995	1204,686	0,603	0,638
Construction	14,527	12,290	1136,727	1275,836	0,475	0,473
Commerce de détail	14,706	12,612	934,347	1191,380	0,666	0,671
Agriculture	15,212	12,601	857,734	960,244	0,666	0,674
Services	15,583	12,748	1083,305	1266,668	0,633	0,665
Habillement	15,645	12,704	1203,872	1360,691	0,597	0,628
Logement	16,910	13,669	783,841	672,754	0,731	0,687
Transport	17,894	13,715	783,603	845,695	0,808	0,728

L'annexe D présente les coefficients de corrélation entre les différentes variables quantitatives. Tous les coefficients de corrélation entre les variables explicatives sont bien en dessous de 0,8, ce qui permet d'exclure toute possibilité de multi-colinéarité (Kennedy, 2008). Les corrélations élevées obtenues entre le langage de risque et extrinsèque, entre le langage de profit et extrinsèque, entre le langage d'intérêt humain et le langage intrinsèque et entre le langage de diversité et le langage intrinsèques sont justifiées parce que le langage de risque, de profit, d'intérêt humain et de diversité sont utilisés dans la construction du langage intrinsèque et extrinsèque. Pour les coefficients entre les variables explicatives et la variable dépendante *Duree*, nous avons une significativité au seuil de 1%. Nous notons des corrélations élevées entre la durée des campagnes et le montant recherché pour l'emprunt mais aussi entre la durée des campagnes et la durée du prêt. Cela soutient l'idée, prouvée dans la littérature, selon laquelle plus le montant recherché est élevé et plus la durée nécessaire pour l'atteindre augmente. Aussi, moins le temps au bout duquel un emprunt sera remboursé est éloigné et plus les prêteurs sont enclins à financer ce projet car ils pourront avoir l'occasion de financer d'autres projets assez tôt avec le même fonds. Contrairement aux corrélations obtenues par Allison et al. (2015), nous obtenons une corrélation négative entre *Interet_hunain* et la durée des campagnes de financement et entre

Intrinseque et la durée des campagnes de financement. En outre, nous notons un lien plus fort entre *Interet_hunain* et la durée des campagnes de financement et entre *Intrinseque* et la durée des campagnes de financement par rapport à Allison et al. (2015).

Les p-values du test d'ANOVA réalisé entre la durée des campagnes de financement des projets et les variables *Secteur*, *Genre* et *Risque* sont inférieures à 5%, indiquant une dépendance entre ces variables de contrôle et la variable dépendante *Duree*.

Pour ce qui est des variables créées à partir des récits de demande de prêts, nous avons examiné leur corrélation avec la durée des campagnes de financement en réalisant une régression MCO. Nous trouvons que 83 sur les 100 coefficients sont statistiquement significatifs pour les groupes et 85 pour les individus. Nous régressons également la variable *Groupe* sur ces 100 variables, et nous obtenons 94 corrélations significatives. Cela indique que les récits de demande de prêts expliquent de façon significative la durée des campagnes ainsi que le traitement.

7. Résultats et discussion

Dans ce chapitre, nous présentons et discutons des résultats de notre étude. Dans la première section de ce chapitre (section 7.1) figurent les résultats relatifs à notre premier objectif, qui porte sur la diversification des portefeuilles de prêts. La deuxième section (section 7.2) comporte les résultats obtenus sur le langage intrinsèque et extrinsèque utilisé dans les récits de demande de prêts. Enfin dans la troisième section (section 7.3), nous détaillons les résultats portant sur les prêts de groupe.

7.1. Principaux résultats sur la diversification des portefeuilles de prêts

De manière spécifique, dans cette section se retrouvent les résultats sur les mesures de diversification et la construction de l'indicateur de diversification (sous-section 7.1.1). Les résultats des deux dernières sous-sections concernent respectivement la classification des portefeuilles en fonction de leur niveau de diversification (sous-section 7.1.2) et l'analyse de leur performance (sous-section 7.1.3).

7.1.1. Résultats des mesures de diversifications et construction de l'indicateur

Le tableau 10 présente des statistiques sur les mesures de diversification. Les valeurs moyennes importantes (0,66 et 0,63) sont obtenues pour le genre de l'emprunteur et le type du porteur de projet, et les faibles pour le secteur et le pays (0,34 et 0,13). Il semble donc que les portefeuilles soient moins diversifiés pour le type et le genre que pour le secteur et le pays. Au-delà du fait que pour le genre et le type, les prêteurs disposent de moins de choix que pour le secteur et le pays, ces différences de niveau de diversification peuvent s'expliquer par les avantages liés à la diversification sectorielle et à la diversification internationale qui ont créé une sorte de contraintes auxquelles tout investisseur se trouve désormais être soumis, c'est-à-dire diversifier le mieux possible son portefeuille en fonction des pays et des secteurs (Emiris 2004; Shepherd, 2010; Christensen, 2018).

Les écart-types obtenus pour le type et le genre sont relativement faibles (0,16 et 0,13) indiquant une faible dispersion autour des valeurs moyennes tandis que les écart-types

obtenus pour secteur et le pays (0,15 et 0,16) laissent entrevoir une importante dispersion autour de leurs valeurs moyennes.

Tableau 10 : Sommaire des mesures de diversification

	Variable	Moyenne	Écart-type	Min	Max	1	2	3	4
1	DIV_Type	0,63	0,13	0,45	1				
2	DIV_Genre	0,66	0,16	0,46	1	0,15**			
3	DIV_Secteur	0,24	0,15	0,11	1	0,19*	0,21*		
4	DIV_Pays	0,13	0,16	0,01	1	0,15**	0,25*	0,07	

Note **:** *** p-value< 0.001, ** p-value< 0.01, * p-value< 0.05, +p-value< 0.1

Nous utilisons, par la suite, ces mesures comme variables pour construire un indicateur synthétique qui permettra d'avoir une idée globale du niveau de diversification des portefeuilles. Le test de Bartlett est significatif (p-value=0,00) et les valeurs de KMO et MSA sont toutes supérieures à 0,5 (tableau 11). Cela nous permet de conclure qu'il est possible d'obtenir une combinaison des quatre mesures en réalisant une ACP.

Tableau 11 : Indices MSA et KMO

	Variables	Libellés	MSA
1	DIV_Type	Indice de diversification pour le type de l'emprunteur	0,63
2	DIV_Genre	Indice de diversification pour le genre de l'emprunteur	0,58
3	DIV_Secteur	Indice de diversification pour le secteur d'activité	0,58
4	DIV_Pays	Indice de diversification pour le pays de l'emprunteur	0,57

Le tableau 12 présente les valeurs propres, le pourcentage d'information restituée et le cumulé du pourcentage d'information restituée par axe. Selon le critère du pourcentage d'information restituée, le nombre minimal d'axes que nous pouvons retenir est 2 et le maximum 3. Le critère de Kaiser-Guttman ne nous permet de retenir que le premier axe.

Tableau 12 : Valeur propre, variance expliquée et cumulée par axe

Axes	Valeur propre	Variance expliquée	Variance expliquée cumulée
Axe 1	1,509	0,377	0,377
Axe 2	0,949	0,237	0,615
Axe 3	0,848	0,212	0,827
Axe 4	0,693	0,173	1

Enfin, le critère du coude de Cattell conduit à deux axes (figure 4). En combinant ces trois critères, nous retenons les deux premiers axes qui restituent 61,5% de l'information.

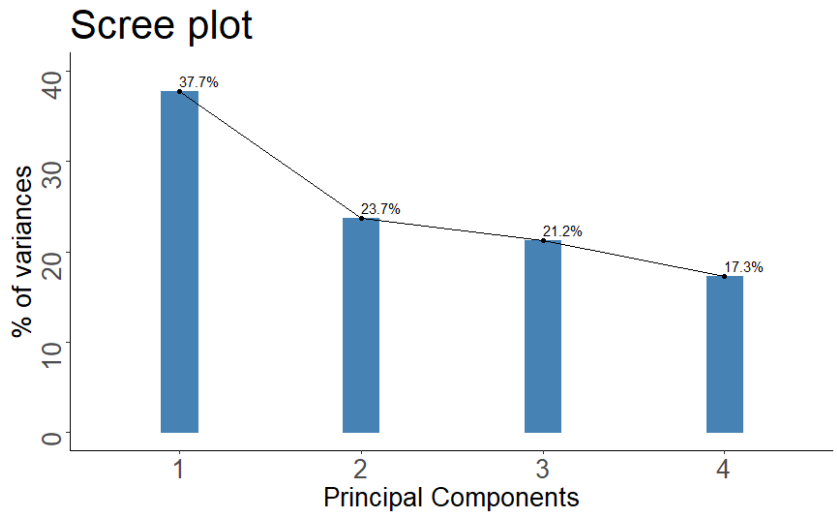

Figure 4 : Représentation des valeurs propres

Notre objectif, en réalisant l'ACP est d'obtenir des pondérations pour chacune des mesures afin de construire l'indicateur synthétique. Le tableau 13 montre les coordonnées et les pondérations de chacune des variables sur chaque axe factoriel. Une augmentation du premier facteur d'une unité entraine une augmentation de 0,586 unité du niveau de diversification pour le type de l'emprunteur, de 0,694 unité pour le genre, de 0,577 unité pour le secteur d'activité et de 0,593 unité pour le pays. L'impact de la variation du premier axe est donc positif pour toutes les variables et plus important pour le genre. Avec le second facteur, une augmentation d'une unité entraine une augmentation de 0,319 unité du niveau de diversification pour le type de l'emprunteur et de 0,614 pour le secteur d'activité mais une baisse de 0,227 unité pour le genre et de 0,647 pour le pays. L'impact de la variation du second axe est donc positif pour le type et le secteur mais négatif pour le genre et le pays. L'impact positif le plus important est obtenu pour le secteur et celui négatif pour le pays. En termes de pondération, DIV_Genre a la pondération positive la plus élevée sur le premier axe mais la plus faible et négative sur le second axe. Sur l'axe 2, DIV_Pays a le poids négatif le plus élevé.

Tableau 13 : Coordonnées des variables par axe

Axes	Coordonnées		Pondérations	
	Axe 1	Axe 2	Axe 1	Axe 2
DIV_Type	0,586	0,319	0,388	0,336
DIV_Genre	0,694	-0,227	0,46	-0,239
DIV_Secteur	0,577	0,614	0,382	0,647
DIV_Pays	0,593	-0,647	0,393	-0,682

A partir de ces pondérations, nous avons construit l'indicateur composite dont les valeurs varient entre -0,038 et 0,112 avec une moyenne de 0,028 (tableau 14). Afin de faciliter son interprétation, nous avons normalisé l'indicateur de sorte que ses valeurs puissent être comprises entre 0 et 1. Des valeurs tendant vers 0 sont synonymes d'une forte diversification tandis que celles tendant vers 1 une forte concentration. Nous présentons dans ce tableau un résumé de l'indicateur non normalisé et celui normalisé. La valeur moyenne de l'indicateur est relativement faible et le maximum est seulement de 0,15. Ceci peut s'expliquer par le fait que les prêteurs que nous étudions ont été sélectionnés sur la base d'un minimum de 10 prêts ; par conséquent ils adoptent déjà une certaine forme de diversification.

Tableau 14 : Statistiques de l'indicateur normalisé et non normalisé

Indicateur	Min	1er Qu.	Médiane	Moyenne	3ème Qu.	Max	Écart-type
Non normalisé	-0,038	0,014	0,0284	0,0281	0,042	0,112	0,025
Normalisé	0	0,052	0,0665	0,0660	0,081	0,151	0,025

7.1.2. Résultats de la classification

La figure 5 indique que la partition optimale avec la classification hiérarchique est une partition en trois classes. La première classe (représentée en rouge) regroupe 32,26% des portefeuilles et la classe 2 (représentée en vert) 49,03%. Enfin, la troisième classe (représentée en bleu) comporte les 18,71% restant (Voir Tableau 15 pour les pourcentages par classe).

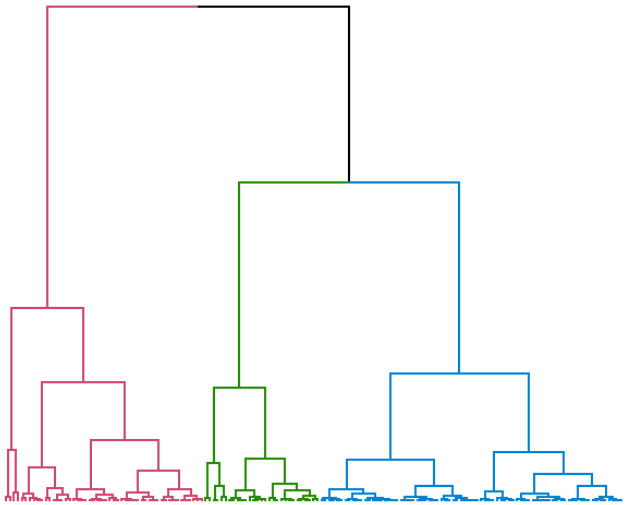

Figure 5: Dendrogramme

La figure 6 nous montre les résultats de la comparaison des 30 indices pour la détermination du nombre optimal de classes avec la méthode k-mean. Comme nous pouvons le voir, la règle de majorité indique une partition en trois classes comme la plus optimale. Le ratio entre la somme des carrées des déviances entre les classes et la déviance totale donne 79,2% indiquant que cette classification est bonne. La première classe est composée de 54,84% des portefeuilles, la deuxième 23,87% et la troisième 21,29%.

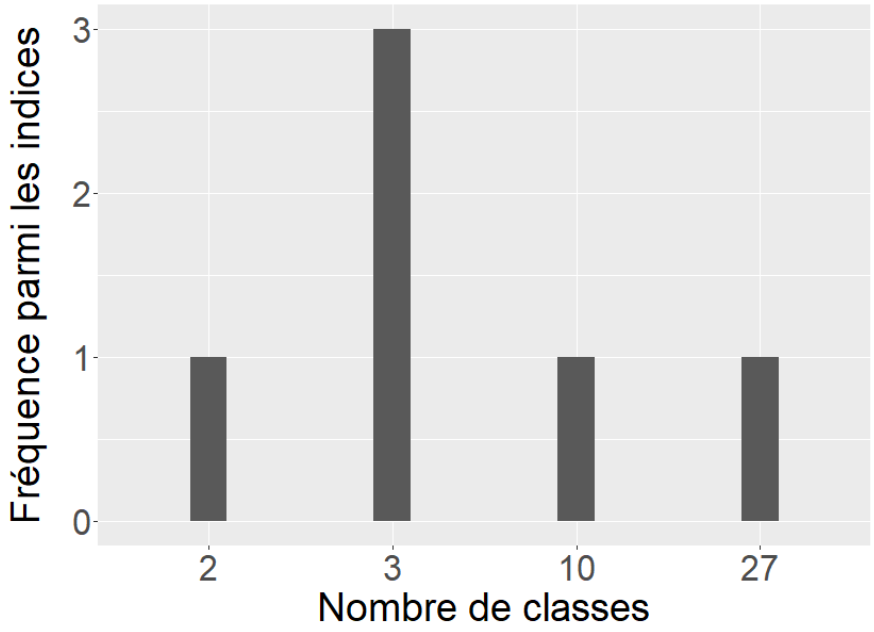

Figure 6 : Résultats des 30 indices (NbClust)

Pour vérifier dans quelle mesure les résultats de la classification hiérarchique et ceux de la méthode k-mean sont concordant, nous calculons le kappa de Cohen pour lequel nous obtenons 82% d'accord. Cela indique un accord parfait entre les résultats de classification hiérarchique et ceux obtenus avec la méthode k-mean. Par conséquent, nous avançons dans notre analyse avec la partition en trois classes obtenues sur la base de la classification non hiérarchique.

Tableau 15 : Résultats de classification

	Classification hiérarchique			Classification kmean		
Classes	N	Portefeuilles	Diversification	N	Portefeuilles	Diversification
Classe 1	50 (32,26%)	1, 9, 11, 13, 17, 19, 20, 21, 27, 33, 35, 37, 41, 105, 53, 118, 57, 61, 62, 65, 69, 73, 74, 75, 77, 79, 83, 85, 93, 99, 50, 117, 55, 121, 123, 125, 127, 129, 131, 133, 137, 138, 141, 145, 151, 153, 49, 113, 54, 119	0,093	37 (23,87%)	1, 35, 79, 129, 9, 41, 85, 133, 11, 49, 105, 137, 13, 54, 113, 141, 17, 55, 117, 145, 19, 57, 119, 151, 20, 62, 121, 153, 21, 65, 123, 27, 69, 125, 33, 73, 127	0,098
Classe 2	76 (49,03%)	2, 42, 89, 134, 18, 64, 106, 38, 3, 43, 90, 135, 22, 66, 107, 39, 4, 45, 91, 139, 25, 67, 108, 86, 5, 46, 94, 142, 26, 70, 109, 87, 6, 51, 95, 143, 29, 71, 110, 130, 7, 52, 97, 146, 30, 72, 111, 132, 8, 56, 98, 147, 31, 78, 114, 10, 58, 100, 149, 32, 80, 115, 14, 59, 101, 150, 34, 81, 122, 15, 63, 103, 154, 36, 82, 126	0,062	85 (54,84%)	3, 22, 42, 61, 77, 91, 106, 126, 146, 4, 25, 43, 63, 78, 93, 107, 130, 147, 5, 26, 45, 64, 80, 94, 108, 131, 149, 6, 29, 46, 66, 81, 95, 109, 132, 150, 7, 30, 50, 67, 82, 97, 110, 134, 154, 8, 31, 51, 70, 83, 98, 111, 135, 10, 36, 52, 71, 86, 99, 114, 138, 14, 37, 53, 72, 87, 100, 115, 139, 15, 38, 58, 74, 89, 101, 118, 142, 18, 39, 59, 75, 90, 103, 122, 143,	0,065
Classe 3	29 (18,71%)	12, 68, 120, 16, 76, 124, 23, 84, 128, 24, 88, 136, 28, 92, 140, 40, 96, 144, 44, 102, 148, 47, 104, 152, 48, 112, 155, 60, 116	0,030	33 (21,29%)	2, 47, 102, 148, 12, 48, 104, 152, 16, 56, 112, 155, 23, 60, 116, 24, 68, 120, 28, 76, 124, 32, 84, 128, 34, 88, 136, 40, 92, 140, 44, 96, 144	0,032
Total	**155**					

7.1.3. Résultats de la performance des classes

Le niveau moyen de diversification des portefeuilles est de 0,098 pour la classe 1, de 0,065 pour la classe 2 et de 0,032 pour la classe 3. Le t-test réalisé pour comparer les niveaux de diversification des classes, prises deux à deux, est significatif (p-value<5%) indiquant que les portefeuilles de classe 3 sont plus diversifiés que ceux des classes 2 et 1. Aussi, cela indique que les portefeuilles de la classe 2 sont plus diversifiés que ceux de la classe 1. La classe 1 est donc celle des portefeuilles à faible diversification, la classe 2 celle des portefeuilles à diversification moyenne et enfin la classe 3 celle des portefeuilles à forte diversification. Ces résultats nous permettent de valider l'hypothèse 1-1 qui stipule que les multi-prêteurs adoptent différentes stratégies de diversification.

Tableau 16 : Niveau de diversification par classe

Classes	Classe 1	Classe 2	Classe 3
DIV_Type	0,718	0,594	0,597
DIV_Genre	0,623	0,668	0,731
DIV_Secteur	0,285	0,224	0,217
DIV_Pays	0,142	0,133	0,092
Indicateur global	0,098	0,065	0,032
% de multi-prêteurs	23,87%	54,84%	21,29%

Compte tenu de ces différences entre les classes, nous nous attendons à observer des différences entre ces trois classes pour ce qui est du pourcentage des campagnes réussies et de leur durée.

Les résultats des tableaux 17, 18 et 19 confirment nos attentes. Ils comportent les résultats du test de proportion et du test de comparaison de moyennes (t-test) réalisés pour comparer le pourcentage de campagnes réussies et la durée moyenne des campagnes réussies entre les paires de classes.

Tableau 17 : Durée moyenne et % de campagnes réussies (classe 1 et classe 2)

Classes	Classe 1	Classe 2	Test de comparaison
% de campagnes réussies	89,92	93,46	Significatif
Durée moyenne des campagnes réussies	17,86	15,72	Significatif

Nous remarquons que les projets des portefeuilles de la classe 1 qui ont réussi représentent 89,92% et ont mis en moyenne 17,86 jours pour se financer. Dans la classe 2, les collectes de fonds qui ont réussi sont de 93,46% avec une durée moyenne de 15,72 jours.

Entre la classe 1 et la classe 2, il y a une différence significative (p-value<5%), non seulement pour le pourcentage de projets totalement financés mais également pour la durée nécessaire à ce financement (2,14 jours). Cela indique que le pourcentage de campagnes réussies dans la classe 2 est plus important que celui de la classe 1. Aussi, la durée des campagnes de la classe 2 qui ont réussi est inférieure à celle des campagnes de la classe 1 qui ont réussi.

Tableau 18 : Durée moyenne et % de campagnes réussies (classe 1 et classe 3)

Classes	Classe 1	Classe 3	Test de comparaison
% de campagnes réussies	89,92	91,15	Significatif
Durée moyenne des campagnes réussies	17,86	17,36	Significatif

La classe 3 affiche un pourcentage de 91,15% et une durée moyenne 17,36 jours. L'écart de 1,23% qui existe entre les pourcentages de campagnes de financement réussies des classes 1 et 3 est significatif (p-value<5%). Par ailleurs, la durée moyenne que requièrent les projets des portefeuilles de la classe 1 pour leur financement et la durée requise par les projets des portefeuilles de la classe 3 sont très proches mais significativement différentes (p-value<5%). Par conséquent, le pourcentage de campagnes réussies dans la classe 3 est plus important que celui de la classe 1 ; et la durée des campagnes de la classe 3 qui ont réussi est inférieure à celle des campagnes de la classe 1 qui ont réussi.

Tableau 19 : Durée moyenne et % de campagnes réussies (classe 2 et classe 3)

Classes	Classe 2	Classe 3	Test de comparaison
% de campagnes réussies	93,46	91,15	Significatif
Durée moyenne des campagnes réussies	15,72	17,36	Significatif

Entre la classe 2 et la classe 3, nous observons également une différence significative (p-value<5%) pour le pourcentage de campagnes de financement ayant réussi (2,31%) et pour leur durée (1,64 jours). Cela montre que le pourcentage des campagnes réussies de la classe 2 est supérieur à celui de la classe 3 et que la durée de ces campagnes dans la classe 2 est en moyenne inférieure à celle des campagnes réussies dans la classe 3.

Ces résultats indiquent qu'un portefeuille avec une faible diversification est moins performant qu'un portefeuille à diversification moyenne et moins performant qu'un portefeuille à diversification forte. Plus encore, il ressort qu'un portefeuille avec une diversification forte est moins performant qu'un portefeuille à diversification moyenne. Par conséquent, une forte et une faible diversification sont coûteuses mais une forte diversification est moins coûteuse qu'une faible diversification.

Les résultats du tableau 20 confirment les précédents résultats car ils montrent que les projets de la classe 1 ($e^{\beta} = 0{,}983, p_{value} < 0{,}01$) et de la classe 3 ($e^{\beta} = 0{,}988, p_{value} < 0{,}01$) se financent moins vite que les projets de la classe 2. Aussi, les projets de la classe 1 se financent encore moins vite que les projets des trois classes.

Nous validons donc l'hypothèse 1-2, qui suggère que le niveau de performance des portefeuilles varie en fonction du niveau de diversification.

Tableau 20 : Résultats de la régression de Cox

Variables	Modèle 1: exp(coef)	Modèle 2 : exp(coef)
Classe=1		0,983***
Classe=3		0,988***
Accomplissement	0,999***	0,999***
Blame	1,004***	1,004***
Nivellement	0,998***	0,998***
Tenacite	0,995***	0,995***
Variete	0,945**	0,942**
Caratere_concret	1***	1***
Preo_actuelle	0,997***	0,997***
Log (Montant_pret)	0,674***	0,674***
Risque_IMF	1,01***	1,01***
Duree_pret	0,943***	0,943***
Couverture=1	0,672***	0,672***
Couverture=2	0,979***	0,979***
P_value (Test du rapport de vraisemblance)	129 492	129 523
Degré de liberté	12	14

Note: *** p-value< 0.001, ** p-value< 0.01, ** p-value< 0.5, +p-value< 0.1

7.2. Principaux résultats sur les récits de demande de prêts

Nous présentons ici, les résultats portant sur les récits de demandes de prêts. Nous débutons avec des résultats qui prouvent la nécessité d'étudier de manière spécifique cette question pour le groupe des multi-prêteurs (sous-section 7.1.1). Par la suite, nous abordons les résultats démontrant un changement dans la manière dont les récits de demandes de prêts affectent les durées des campagnes de financement (sous-section 7.1.2). La sous-section 7.2.3 est consacrée aux les résultats obtenus, avec l'échantillon des multi-prêteurs. Enfin, la sous-section 7.2.4 s'intéresse à la sensibilité de nos résultats.

7.2.1. Résultats sur l'intérêt à étudier les récits de demandes de prêts des multi-prêteurs

Dans le tableau 21, nous avons reporté les résultats de la régression des moindres carrés ordinaires. Le modèle 1 est effectué en utilisant uniquement les variables de contrôle. Dans le modèle 2, nous incluons en plus des variables de contrôles, la variable *Preteur*. En faisant cette modélisation, nous voulons vérifier si la différence significative de durée de campagnes observée entre l'échantillon Éch2b et l'échantillon Éch3b demeure toujours lorsque nous contrôlons pour d'autres caractéristiques. Le résultat obtenu indique qu'entre la durée moyenne des campagnes de financement auxquelles les multi-prêteurs ont contribué (Échantillon Éch2b) et celle des campagnes auxquelles ils n'ont pas participé (Échantillon Éch3b), il existe une différence belle et bien significative. Plus précisément, la présence des multi-prêteurs en tant que prêteurs dans les campagnes de financement est associée à une augmentation de la durée des campagnes (β=0,293; $p<0,1\%$). Ce résultat souligne, une fois de plus, l'intérêt qu'il y a à étudier les choix de financement des multi-prêteurs.

Un autre point sur lequel il est possible que les choix des multi-prêteurs se démarquent de ceux des autres prêteurs, réside dans la manière dont le langage intrinsèque et le langage extrinsèque affectent la durée des campagnes de financement des projets qu'ils choisissent. Pour vérifier si une telle différence est bien réelle, nous regardons à la significativité du coefficient de *Extrinseque x Preteur* et du coefficient de *Intrinseque x Preteur*. Si ces coefficients sont significatifs, alors nous pouvons conclure que les multi-prêteurs se distinguent des autres prêteurs Kiva de par la manière dont le langage intrinsèque et le

langage extrinsèque agissent sur leurs décisions de financement. Toutefois, avant de procéder à ces vérifications, il nous faut nous assurer que le langage intrinsèque et le langage extrinsèque utilisés dans les récits de demande de prêts affectent l'attractivité des prêteurs comme le suggèrent les études antérieures (Allison et al., 2015). Les résultats (Modèle 3, Tableau 21) indiquent l'existence d'un lien significatif seulement entre le niveau de langage intrinsèque et la durée des campagnes de financement. Nous trouvons, en effet, qu'un niveau élevé de langage intrinsèque est associé à une durée plus longue de campagnes (β=0,003; p<0,1%) et que le niveau de langage extrinsèquement n'est nullement associé à la durée des campagnes de financement (β=-0,0005; p=0,50). Par conséquent, nous allons nous intéresser uniquement à comment l'association du niveau de langage intrinsèque à la durée des campagnes varie en fonction de la présence des multi-prêteurs.

À partir des résultats du modèle 4, nous voyons que le coefficient de la variable d'interaction *Intrinseque x Preteur* est significatif (β=0,012; p<0,1%), ce qui indique que l'association entre le niveau du langage intrinsèque et la durée des campagnes varie en fonction de la présence ou non des multi-prêteurs. En fait, nous obtenons une différence de 1,23% entre l'impact du langage intrinsèque sur la durée des campagnes auxquelles les multi-prêteurs ont contribué et celui du langage intrinsèque sur la durée des campagnes auxquelles les multi-prêteurs n'ont pas contribué. Cela confirme que les multi-prêteurs sont un groupe de prêteurs qui se distinguent des autres prêteurs, et étudier leurs choix de prêts nous permettrait de mieux compréhension les facteurs qui sous-tendent leur décision.

Tableau 21 : Résultats de la régression MCO avec les effets d'interaction

Variables	Modèle 1	Modèle 2	Modèle 3	Modèle 4
Intrinseque			0,003***(0,001)	-0,006***(0,001)
Extrinseque			-0,0005(0,001)	0,001(0,001)
Preteur=1		0,293***(0,002)		0,294***(0,002)
Intrinseque x Preteur				0,013***(0,001)
Log (Montant_pret)	0,433***(0,002)	0,384***(0,002)	0,433***(0,001)	0,384***(0,002)
Risque_IMF	0,004***(0,001)	0(0,001)	0,004***(0,001)	0(0,001)
Duree_pret	0,028***(0)	0,029***(0)	0,028***(0)	0,029***(0)
Accomplissement	0+(0)	0(0)	0(0)	0(0)
Blame	-0,003***(0)	-0,003***(0)	-0,003***(0)	-0,003***(0)
Nivellement	-0,004***(0)	-0,004***(0)	-0,004***(0)	-0,004***(0)
Tenacite	0,002***(0)	0,002***(0)	0,002***(0)	0,002***(0)
Variete	0,22***(0,017)	0,248***(0,017)	0,248***(0,017)	0,248***(0,017)
Caractere_concret	0,001***(0)	0,001***(0)	0,001***(0)	0,001***(0)
Preo_actuelle	-0,001***(0)	-0,001***(0)	-0,001***(0)	-0,001***(0)
Couverture				
Secteur				
Pays				
Constante	-2,031***(0,041)	-1,833***(0,041)	-2,030***(0,041)	-1,835***(0,041)
R^2	0,2282	0,2392	0,2282	0,2393
DDL	111	112	113	116
DDL (résidus)	1 294 213	1 294 212	1 294 211	1 294 208

*** p-value< 0.001, ** p-value< 0.01, ** p-value< 0.5, +p-value< 0.1

7.2.2. Résultats sur le changement dans le comportement de prêts des prêteurs Kiva

Les résultats du modèle 3 (Tableau 21) diffèrent de ce qui a été observé par le passé, notamment par Allison et al. (2015). Ces auteurs ont, en fait, montré qu'un niveau élevé de langage intrinsèque est associé une réduction de la durée des campagnes et qu'un niveau élevé de langage extrinsèque est associé à une augmentation de la durée des campagnes. Se serait-il produit, avec le temps, un changement dans la manière dont ces facteurs affectent l'attractivité des prêteurs Kiva? c'est la question que nous nous sommes posée et pour en avoir le cœur net, nous avons vérifié s'il existe un sous-échantillon pour lequel les résultats de Allison et al. (2015) sont vérifiables.

Le tableau 22 présente les résultats de cette investigation. Les modèles 1 et 2 de ce tableau portent sur un sous-échantillon allant de 2006 à 2009. Les résultats du modèle 2 montrent qu'effectivement et comme dans le cas de Allison et al. (2015), un niveau élevé de langage intrinsèque est associé à une courte durée de campagne (β=-0,024; $p<0,1\%$), et un niveau élevé de langage extrinsèque est associé à une durée de compagne plus longue (β=0,013; $p<0,1\%$).

Dans les modèles 3 et 4, nous utilisons le sous-échantillon 2010-2019. Avec les résultats du modèle 4, nous observons un renversement, c'est-à-dire qu'un niveau élevé de langage intrinsèque est associé à une durée des campagnes de financement plus longue (β=0,009; $p<0,1\%$) et qu'un niveau élevé de langage extrinsèque est associé à une durée plus courte (β=-0,004; $p<0,1\%$).

Il y a, par conséquent, eu un changement dans la manière dont les facteurs intrinsèques et extrinsèques contenus dans les récits de demande de prêts influent sur la durée des campagnes de financement. Pour les variables Tenacite et Preo_actuelle, nous observons également un changement entre les deux échantillons. En fait, avant 2010, ces deux variables étaient associées à une réduction de la durée des campagnes de financement mais à partir de 2010 elles sont associées à une augmentation de cette durée.

Selon Allison et al. (2013), faire preuve de ténacité et aborder des préoccupions d'actualité inspirent de la confiance aux prêteurs et révèlent l'importance de l'action que l'emprunteur compte menée, ce qui devrait avoir pour conséquence une réduction de la durée des campagnes. Cette assertion est vérifiée pour l'échantillon 2006-2009 mais pas pour celui

2010-2019. Une potentielle explication au renversement de tendance observé pour le langage intrinsèque et celui extrinsèque pourrait donc être liée au changement observé pour ces deux variables de contrôle.

Pour tester nos hypothèses 2-1 et 2-2 dans la suite, nous allons donc prendre en compte ces sous-périodes.

Tableau 22 : Résultats de la régression MCO de 2006 à 2009 et de 2010 à 2019

Variables	2006-2009		2010-2019	
	Modèle 1	Modèle 2	Modèle 3	Modèle 4
Variables indépendantes				
Intrinseque		-0,024***(0,002)		0,009***(0,001)
Extrinseque		0,013***(0,003)		-0,004***(0,001)
Variables de contrôle				
Log (Montant_pret)	0,461***(0,006)	0,461***(0,002)	0,446***(0,001)	0,466***(0,001)
Risque_IMF	-0,007*(0,003)	-0,008*(0,003)	-0,005***(0,001)	-0,005(0,001)
Duree_pret	0,022***(0,001)	0,022***(0,001)	0,028***(0)	0,027***(0)
Accomplissement	0,002***(0)	0,002***(0,0003)	0(0)	0(0)
Blame	-0,041***(0,002)	-0,042***(0,002)	0(0)	-0,0001***(0)
Nivellement	-0,006***(0,001)	-0,005***(0,001)	-0,003***(0)	-0,003***(0)
Tenacite	-0,001***(0)	-0***(0)	0,003***(0)	0,003***(0)
Variete	0,186***(0,053)	0,211***(0,053)	0,229***(0,018)	0,232***(0,017)
Caractere_concret	0,002***(0)	0,003***(0)	0***(0)	0,001***(0)
Preo_actuelle	-0,005***(0)	-0,005***(0)	0,001***(0)	0,001***(0)
Couverture				
Secteur				
Pays				
Constante	-2,031***(0,041)	-1,859***(0,069)	-3,049***(0,082)	-3,059***(0,041)
R^2	0,2094	0,21	0,2139	0,2141
DDL	73	75	109	111
DDL (résidus)	149 772	149 770	1 144 369	1 144 367

Note: *** p-value< 0.001, ** p-value< 0.01, ** p-value< 0.5, +p-value< 0.1

7.2.3. Résultats du langage intrinsèque et extrinsèque pour l'échantillon des multi-prêteurs

Nous vérifions les hypothèses 2-1 et 2-2 sur les périodes 2006-2019, 2006-2009 et 2010-2019. Les résultats sont présentés dans le tableau 23. Pour ces trois périodes, nous avons les modèles 1, 3 et 5 qui contiennent uniquement les variables de contrôle et les modèles 2, 4 et 6 qui comportent, en plus des variables de contrôle, les variables mesurant le niveau de langage intrinsèque et le niveau de langage extrinsèque dans les récits.

Les résultats du modèle 2 nous amènent à invalider les hypothèses 2-1 et 2-2. Selon l'hypothèse 2-1, un important niveau de langage intrinsèque devrait être associé à une courte durée de campagne, mais nous observons qu'une augmentation de 1% du score de langage intrinsèque est associée à une augmentation du log de la durée des campagnes d'environ 0,3%. Quant à l'hypothèse 2-2, elle stipule qu'un important niveau de langage extrinsèque devrait être associé à une longue durée de campagne, mais nous trouvons qu'une augmentation de 1% du score de langage extrinsèque est associée à une réduction du log de la durée des campagnes d'environ 0,4%.

Avec les modèles 4 et 6, nous essayons de voir ce qu'il en est pour les deux sous-périodes identifiées dans la section 7.2.2. Les résultats nous montrent, qu'avant 2010, une augmentation de 1% du score du langage intrinsèque et une augmentation de 1% du score du langage extrinsèque sont respectivement associées à une réduction de 3,4% du log de la durée des campagnes et à une augmentation de 2% du log de la durée des campagnes. Les hypothèses 2-1 et 2-2 sont donc valides sur cette sous-période. Sur la sous-période 2010-2019, une augmentation de 1% du score du langage intrinsèque et une augmentation de 1% du score du langage extrinsèque sont respectivement associées à une augmentation de 0,8% du log de la durée des campagnes et à une réduction de 0,6% du log de la durée des campagnes. Les hypothèses 2-1 et 2-2 ne sont donc pas valides sur cette sous-période.

Le changement dans les comportements de prêts que ces résultats mettent en lumière a également été observé par Priberny et Sommer (2021). Leur étude montre que le comportement de prêts des prêteurs Kiva en ce qui concerne les prêts à usage personnel a changé après la crise de 2009. Comme explication, les auteurs ont soulevé le fait que le principal objectif des prêteurs Kiva est de contribuer à la réduction de la pauvreté. Cette explication peut servir également dans notre cas. En effet, le score de langage extrinsèque

indique le niveau de risque du projet et le profit qui est attendu. Une valeur élevée pour le score de langage extrinsèque sous-entend donc plus de risque mais également un revenu futur plus élevé. En choisissant de tels projets, il se pourrait que les prêteurs aient l'impression de contribuer efficacement à la réduction de la pauvreté car il s'agit de projets rentables dont les retombés pourraient énormément contribuer à l'amélioration des conditions de vie des bénéficiaires. Quant au langage intrinsèque, une valeur élevée du score indique que l'emprunteur met un accent important sur la diversité et l'intérêt humain. Selon Larrimore et al. (2011), les prêteurs à la recherche de projets rentables considèrent que les emprunteurs qui mettent l'accent sur les informations relatives à la famille, aux amis ou encore à la religion pratiquée comme des mauvais emprunteurs. La tendance des prêteurs qui semble se porter maintenant vers des projets rentables est, par conséquent, cohérente avec le fait que le langage intrinsèque soit positivement corrélé avec la durée des campagnes.

Tableau 23 : Résultats de MCO pour l'échantillon des multi-prêteurs (Éch2b)

	2006-2019		2006-2009		2010-2019	
Variables	Modèle 1	Modèle 2	Modèle 3	Modèle 4	Modèle 5	Modèle 6
Variables indépendantes						
Intrinseque		0,003***(0,001)		-0,034***(0,005)		0,008***(0,001)
Extrinseque		-0,004***(0,001)		0,020***(0,006)		-0,006***(0,001)
Variables de contrôle						
Log (Montant_pret)	0,401***(0,002)	0,401***(0,002)	0,433***(0,011)	0,432***(0,011)	0,410***(0,002)	0,410***(0,002)
Risque_IMF	-0,018***(0,001)	-0,018***(0,002)	0,031***(0,006)	0,030***(0,006)	-0,025***(0,001)	-0,025(0,001)
Duree_pret	0,035***(0)	0,035***(0,001)	0,042***(0,002)	0,042***(0,002)	0,034***(0)	0,034***(0)
Accomplissement	0+(0)	0(0)	-0,001(0,001)	-0(0)	0(0)	0(0)
Blame	0,001(0,001)	0,001(0,001)	-0,042***(0,004)	-0,043***(0,004)	0,002***(0)	0,002***(0)
Nivellement	-0,004***(0)	-0,004***(0)	-0,007***(0,002)	-0,006***(0,001)	-0,003***(0)	-0,003***(0)
Tenacite	0,002***(0)	0,002***(0)	0(0,001)	0,001(0,006)	0,003***(0)	0,003***(0)
Variete	0,214***(0,021)	0,216***(0,021)	-0,078***(0,108)	-0,045(0,108)	0,257***(0,021)	0,262***(0,021)
Caractere_concret	0*(0)	0+(0)	0,001*(0)	0,002**(0)	0**(0)	0 **(0)
Preo_actuelle	0(0)	0(0)	-0,004***(0,001)	-0,003***(0,001)	0,001***(0)	0,001***(0)
Couverture						
Secteur						
Pays						
Constante	-1,929***(0,073)	-1,931***(0,073)	-1,684***(0,143)	-1,759***(0,144)	-2,815***(0,087)	-2,825***(0,121)
R^2	0,2187	0,2188	0,2438	0,2451	0,2162	0,2163
DDL	110	112	73	75	108	110
DDL (résidus)	787 078	787 076	33 878	33 876	753 128	753 126

7.2.4. Mesure de la sensibilité des résultats MCO

Notre objectif ici est de vérifier, en changeant de méthode, à quels points nos résultats sont sensibles. Nous utilisons pour cela le modèle de Cox dont les résultats sont récapitulés dans le tableau 24 à côté de ceux obtenus avec la régression des moindres carrés. Nous y présentons le signe des coefficients significatifs obtenus avec la régression des moindres carrées ordinaires et le modèle de Cox. Un coefficient négatif dans le modèle de Cox indique que la variable concernée est associée à une réduction de la probabilité de réussite des campagnes de financement tandis qu'un coefficient positif indique que la variable est associée à une augmentation de la probabilité de réussite des campagnes de financement.

Les résultats changent à trois niveaux (surlignés en gras). Le coefficient du langage intrinsèque est non significatif pour l'échantillon des multi-prêteurs (Éch2b, 2006-2019) avec le modèle de Cox alors qu'il l'est dans la régression des moindres carrés ordinaires. Néanmoins, il est significatif et concordant dans le cas des deux sous-périodes (Éch2b, 2006-2009 et 2010-2019), confirmant nos conclusions précédentes.

Pour ce qui est des prêts auxquels les multi-prêteurs n'ont pas contribué (Éch3b, 2006-2019 et 2010-2019), le coefficient du langage extrinsèque est significatif avec le modèle de Cox mais il ne l'est pas avec la régression des moindres carrés ordinaires.

Toutefois, le fait qu'un niveau élevé de langage extrinsèque soit associé à une probabilité de réussite de campagne plus élevée dans l'échantillon Éch2b mais à une probabilité plus faible dans l'échantillon Éch3b est concordant avec notre conclusion selon laquelle les multi-prêteurs préfèrent soutenir des projets rentables afin de s'assurer de contribuer efficacement à la réduction de la pauvreté dans les pays en voie de développement.

Tableau 24 : Récapitulatif des résultats MCO et COX

		Langage intrinsèque		Langage extrinsèque	
Échantillon		MCO	COX	MCO	COX
Éch1b	2006-2009	Négatif	Positif	Positif	Négatif
	2010-2019	Positif	Négatif	Négatif	Positif
	2006-2019	Positif	Négatif	Non significatif	Non significatif
Éch2b	2006-2009	Négatif	Positif	Positif	Négatif
	2010-2019	Positif	Négatif	Négatif	Positif
	2006-2019	**Positif**	**Non significatif**	Négatif	Positif
Éch3b	2006-2009	Négatif	Positif	Positif	Négatif
	2010-2019	Positif	Négatif	**Non significatif**	**Positif**
	2006-2019	Non significatif	Non significatif	**Non significatif**	**Négatif**

7.3. Principaux résultats sur les prêts de groupe

Dans cette section, nous présentons et discutons des résultats obtenus pour les prêts de groupe à la lumière des résultats d'études antérieures. Nous débutons par les mesures d'évaluation de la performance des méthodes mises en œuvre et terminons par les effets estimés obtenus.

7.3.1. Analyse des résultats de métriques d'évaluation

Le tableau 25 montre les résultats des mesures de performances pour les différentes méthodes. Plus la précision et le score F1 augmentent et meilleures sont les prédictions. Avec la racine de l'erreur quadratique moyenne (REQM), les prédictions sont meilleures lorsque cette mesure est faible.

Tableau 25 : Résultats des métriques d'évaluation

Méthode	Score de propension		Groupe (REQM)	Individu (REQM)
	F1	Précision		
En absence des récits				
ElasticNet	0,568	0,899	11,339	11,52
Forêt aléatoire	0,730	0,902	11,568	10,534
En présence des récits				
ElasticNet	0,842	0,927	11,072	10,870
Forêt aléatoire	0,860	0,966	10,605	10,475
Perceptron multicouche	0,855	0,951	10,51	9,951
MLCT profond	0,961	0,984	9,453	10,072

D'une manière générale, nous remarquons que les prédictions s'améliorent lorsque les données issues des récits de demande de prêts sont incluses dans l'analyse car avec ces données les précisions et les scores F1 sont plus élevés tandis que les REQM sont plus faibles. Cette amélioration est surtout notable pour le score de propension. Cela prouve la nécessité de tenir compte des récits de demandes de prêts.

Comparer aux métriques d'évaluation des autres méthodes, celles de la méthode de régression pénalisée sont les moins bonnes lorsqu'on est en présence des récits de demandes de prêts. Pour le score de propension, elles sont les plus basses et pour les résultats de traitement, elles sont les plus élevées. En revanche, la méthode de forêts

aléatoires donne de moins bonnes prédictions que la méthode de régression pénalisée pour les résultats de traitement en absence des récits.

Le MLCT profond donne de meilleures prédictions que les autres méthodes sauf dans le cas des résultats de traitement des emprunteurs seuls où le REQM du perceptron multicouche est plus petit que le sien (REQM=9,95 et REQM=10,07). La méthode de forêts aléatoires donne de meilleures prédictions que le perceptron multicouche pour le score de propension. Dans notre cas, et contrairement au cas de Pham et Shen (2016) où le MLCT profond a donné de meilleures prédictions en tout point, aucune des méthodes ne supplante les autres en tout point. Cela indique l'intérêt d'utiliser l'estimateur moyen de Antonelli et Cefalu.

7.3.2. Analyse des résultats de l'effet moyen de traitement

Le tableau 26 comporte les résultats d'estimation de l'effet moyen de traitement (EMT) pour chacune des méthodes. En dehors de la naïve, nous avons deux effets estimés pour chaque méthode : un premier qui correspond au cas où nous ne tenons pas compte des récits de demande de prêts et un second pour le cas où ils sont pris en compte.

Tableau 26 : Résultats des méthodes

Méthode	Effet moyen	Erreur standard
En absence des récits de demande de prêts		
Naïve	1,05	0,046
ElasticNet	-6,053	0,08
Estimateur de double sélection		
Lasso	-3,619	0,036
Estimateur doublement robuste		
ElasticNet	-6,043	0,315
Forêt aléatoire	-5,702	0,541
Estimateur ciblé du maximum de vraisemblance		
ElasticNet	-6,581	0,371
Forêt aléatoire	-5,716	0,505
En présence des récits de demande de prêts		
ElasticNet	-6,107	0,07
Estimateur de double sélection		
Lasso	-2,942	0,047
Estimateur doublement robuste		
ElasticNet	-8,113	1,896
Forêt aléatoire	-5,788	0,075
Perceptron multicouche	-4,952	1,849
MLCT profond	-0,894	0,811
Estimateur ciblé du maximum de vraisemblance		
ElasticNet	-8,275	1,862
Forêt aléatoire	-5,790	0,077
Perceptron multicouche	-5,003	1,865
MLCT profond	-0,915	0,881
Estimateur moyen	-4,338	0,941

La méthode naïve donne un effet moyen de 1,05 jours avec un écart-type de 0,05. Ce résultat indique que la constitution de groupe augmente le temps de financement des prêts et qu'il est préférable d'être seul que d'être un groupe. Comme souligné dans la méthodologie, cette méthode est susceptible de conduire à un résultat biaisé dans le cas des données observationnelles. C'est ce que démontrent les résultats que nous obtenons avec les autres méthodes. En effet, en dehors de cette méthode, toutes les autres donnent des résultats qui indiquent que la constitution en groupe des emprunteurs réduit le temps de financement.

La régression linéaire pénalisée donne un effet estimé quasiment identique en présence (EMT=-6,05 et ES=0,08) et en absence des récits (EMT=-6,11 et ES =0,07). Cela donne l'impression que cette méthode n'est pas adaptée pour traiter des données importantes. En absence des récits, la méthode de double sélection donne l'effet le plus bas en valeur absolue (EMT=-3,62 et ES=0,04). En présence des récits, le résultat obtenu avec cette méthode indique qu'être un groupe d'emprunteurs plutôt qu'un emprunteur seul réduit la durée des campagnes de 2,94 jours.

Que ce soit en absence ou en présence des récits de demande de prêts, l'estimateur doublement robuste et l'estimateur ciblé du maximum de vraisemblance donnent dans résultats très similaires. En fait, en absence des récits les résultats de ces deux estimateurs sont très proches pour la régression pénalisée (respectivement EMT=-6,04, ES=0,31 et EMT=-6,58, ES=0,37) et pour la méthode des forêts aléatoires (respectivement EMT= -5,70, ES=0,54 et EMT=-5,72, ES=0,50). De même, en présence des récits de demande de prêts, leurs résultats sont très proches : régression pénalisée ElasticNet (respectivement EMT= -8,11, ES=1,90 et EMT=-8,27, ES=1,86), forêts aléatoires (respectivement EMT=-5,79, ES=0,07 et EMT=-5,79, ES=0,08), perceptron multicouche (respectivement EMT=-4,95, ES=1,85 et EMT=-5, ES=1,86) et MLCT profond (respectivement EMT= -0,89, ES=0,81 et EMT=-0,91 ES=0,88). Il est possible que cela soit lié à la propriété de double robustesse que possèdent ces deux estimateurs.

Par rapport à l'étude de Pham et Shen (2015), nous obtenons pour la plupart des méthodes des effets plus importants en valeur absolue sauf pour l'estimateur doublement robuste (EMT= -0,89, ES=0,81) et l'estimateur ciblé du maximum de vraisemblance (EMT=-5,79,

ES=0,07; EMT=-5, ES=1,86 et EMT= -0,91, ES=0,88). En outre, nous obtenons des effets négatifs avec les régressions pénalisées ce qui n'est pas le cas pour Pham et Shen (2015).
L'estimateur moyen donne un effet estimé de -4,34, ce qui veut dire que la constitution des emprunteurs en groupe réduit le temps de financement de 4,34 jours. Il y a, par conséquent, un gain en termes de temps de financement pour les prêts de groupe par rapport aux prêts individuels. Notre hypothèse 3 est pour cela validée.

Plusieurs auteurs, en microcrédit, sont parvenus à la même conclusion que nous qu'ils justifient par le taux de remboursement élevé associé aux prêts de groupe (Ghatak, 1999; Stiglitz, 1990; Besley et Coates, 1995; Bratton, 1986; Paxton, 1996; Zeller, 1998). En effet, les groupes sont essentiellement composés d'emprunteurs qui se connaissent et qui mutuellement se surveillent et exercent des pressions les uns sur les autres. Dans l'environnement de prêts pair-à-pair, certains auteurs ont également montré que les prêteurs préfèrent les prêts de groupe pour deux raisons : le désir de se faire rembourser afin de continuer de prêter à d'autres et le désir de maximiser l'impact de chaque contribution. Sur Kiva, les prêteurs sont libres d'investir dans autant de projets qu'ils souhaitent et à un montant aussi minime que 25 $ US. Ils sont donc en mesure de diversifier convenablement leurs portefeuilles de prêts afin de minimiser les risques de perte. Une explication valable surtout dans notre cas, puisque nous étudions les prêts des multi-prêteurs, est le désir de maximiser l'impact de leur aide. Selon Andreoni (2007), la valeur perçue de l'altruisme par le donateur augmente avec le nombre de bénéficiaires. Ainsi, les prêteurs Kiva en prêtant à un groupe plutôt qu'à un seul individu ont le sentiment que leur contribution a un impact plus grand. Cette explication permet également de mieux comprendre le fait que l'effet causal que nous obtenons soit plus important que celui obtenu par Pham et Shen (2015).
En revanche, Ly et Mason (2012b) et en particulier Galak et al. (2011) mentionnent l'existence d'un effet de victime identifiante chez les prêteurs qui les amène à préférer les emprunteurs seuls au groupe. Selon ces auteurs, un emprunteur seul dans le besoin suscite plus d'émotion qu'un groupe d'emprunteur. Galak et al. (2011) montre, par ailleurs, que les prêteurs aiment prêter aux emprunteurs qui leur ressemblance, notamment les emprunteurs dont le prénom débute par la même lettre qu'eux, les emprunteurs qui exercent la même profession qu'eux ou encore les emprunteurs qui sont du même genre qu'eux. Selon ces résultats, il serait sans doute difficile pour un prêteur d'établir un tel lien avec un groupe

d'emprunteurs. L'une des raisons possibles à la différence entre les résultats de ces auteurs et nos résultats est la différence dans les données. Il s'agit de différence en termes de volume de données. Ly et Mason (2012b) ont utilisé des données qui s'étalent sur 3 ans tandis que l'échantillon de Galak et al. (2011) ne couvrent que 8 mois. Une autre différence dans les données se situe au niveau des variables explicatives. Contrairement à ces deux auteurs, nous utilisons les récits de demande de prêts. L'importance de ces récits dans le processus de financement des prêts a été largement documentée (Moysidou, 2016 ; Ordanini et al., 2011 ; Schwienbacher et Larralde, 2012, Allison et al., 2013; Herzenstein et al., 2011; Gao et Lin, 2015). La seconde raison que nous pouvons avancer réside au niveau de la méthodologie. Par exemple, l'étude de Galak et al. (2011) démontre, toutes choses étant égales, qu'Il existe un lien positif entre le nombre d'emprunteurs et le temps de financement d'une part et entre le nombre d'emprunteurs et le montant en $ US qu'un prêteur est susceptible de prêter. De notre côté, nous montrons plutôt, toutes choses étant égales, que le fait d'être un groupe occasionne une réduction du temps de financement. En d'autres mots, notre étude identifie, sous couvert des hypothèses de nos modèles, l'effet que produit sur la durée des campagnes de financement, le fait d'être un groupe.

Conclusion

Ce mémoire porte sur le financement participatif et répond à trois objectifs principaux. Le premier consiste à examiner les stratégies de diversification qu'adoptent les multi-prêteurs Kiva. Par multi-prêteurs nous entendons des prêteurs qui ont soutenu au moins 10 projets. Selon Lin et Boh (2019), le rôle de ces prêteurs sur les plateformes de financement participatif est capital car ils ont tendance à soutenir les projets au début de leur campagne, suscitant ainsi l'intérêt d'autres prêteurs pour ces projets. Le deuxième objectif est d'analyser le lien entre le langage utilisé dans les récits de demande de prêts et la durée des campagnes de financement des projets sélectionnés par les multi-prêteurs. Enfin, le troisième objectif est d'évaluer l'effet d'être un groupe d'emprunteurs plutôt qu'un emprunteur seul sur la durée des campagnes de financement des projets que sélectionnent les multi-prêteurs.

Pour atteindre le premier objectif, nous avons examiné comment les multi-prêteurs diversifient leur portefeuille de prêts en fonction du pays des emprunteurs, leur genre, leur secteur d'activité et leur type. A cet effet, nous utilisons un échantillon de 155 portefeuilles composés de 891 171 projets publiés sur Kiva entre octobre 2006 et janvier 2019. Nous avons construit un indicateur mesurant le niveau de diversification des portefeuilles, créé des classes de portefeuilles en fonction des niveaux de diversification et exécuté un modèle de Cox afin d'examiner la différence de performance qu'il y a entre les classes de portefeuilles identifiées. Les résultats obtenus révèlent trois niveaux de diversification (faible, moyenne et élevé) et également qu'une diversification moyenne permet au prêteur de choisir des projets qui réussissent leur campagne de collecte de fonds en peu de temps. La majorité des multi-prêteurs étudiés adoptent une diversification moyenne.

Le deuxième objectif a été atteint en étudiant le lien entre le langage intrinsèque et la durée des campagnes d'une part et le lien entre le langage extrinsèque et la durée des campagnes d'autre part. Nous utilisons, pour cela, un échantillon de 787 189 projets postés sur Kiva entre octobre 2006 et janvier 2019 et adoptons la même méthodologie que Allison et al. (2015). Alors que nous nous attendions à obtenir les mêmes résultats que ces auteurs, nous parvenons à des résultats contraires. En effet, le langage intrinsèque s'est révélé être positivement associé à la durée des campagnes et le langage extrinsèque négativement. En

revanche, pour les 507 136 projets postés entre octobre 2006 à janvier 2009 et qui n'ont pas bénéficié de la contribution des multi-prêteurs, les résultats indiquent une absence totale de lien entre la durée des campagnes et le langage intrinsèque ainsi que le langage extrinsèque. Par conséquent, il semble que les multi-prêteurs voient au-travers des projets présentés comme rentables un moyen d'être plus utile dans la lutte contre la pauvreté.

Par ailleurs, nos résultats mettent en évidence un changement dans le comportement de prêts des prêteurs Kiva au fil du temps. En effet, nous avons remarqué, qu'avant 2010 le langage intrinsèque était associé à une réduction du temps nécessaire au financement complet des projets et que le langage extrinsèque était associé à une augmentation de ce temps mais qu'à partir de 2010, cela s'est inversé. Nous aboutissons à ces mêmes conclusions lorsque nous considérons uniquement l'échantillon des multi-prêteurs. Priberny et Sommer (2021) ont aussi remarqué, qu'après la crise de 2009, les prêts à usage personnels se finançaient plus souvent et plus vite qu'avant la crise.

Enfin, pour le troisième objectif nous montrons, avec un échantillon de 898 772 projets, qu'être un groupe d'emprunteurs plutôt qu'emprunteur seul réduit le temps nécessaire jusqu'au financement complet des projets. Étant donné que sur Kiva, les prêteurs ne perçoivent pas d'intérêt sur leur investissement, ce résultat suggère que les multi-prêteurs préfèrent les groupes aux emprunteurs seuls car ils souhaitent accroître l'impact de leur aide. Ce résultat peut être vu comme un complément des résultats obtenus avec l'objectif 2 qui prouvent que les projets rentables se financent désormais plus vite, confirmant que les multi-prêteurs visent réellement la réduction de la pauvreté au-travers de leur contribution. En termes de méthodologie, nous avons apporté un plus à ce qu'a fait Pham et Shen (2016) en utilisant l'estimateur moyen de Antonelli et Cefalu et en recourant à un algorithme génétique pour le choix des paramètres de nos réseaux de neurones.

La principale limite de cette étude est le biais de sélection, plus précisément le biais des survivants, engendré par la manière dont nous sélectionnons les multi-prêteurs et par le fait que nous ne considérons que les projets qui ont réussi leur campagne puisqu'il n'est pas possible d'observer la durée des campagnes de financement pour les projets dont la campagne n'a pas réussi. Ce biais a tendance à conduire à des estimations biaisées soit vers le haut ou vers le bas, dépendamment de l'importance des informations omises dans l'analyse (Heckman 1979; Boehmke et al., 2006). Il est également possible que ce biais

affecte la significativité des estimés (Storey 2000). Une façon de prendre en compte ce biais aurait été d'utiliser l'approche d'estimation en deux étapes de Heckman. Nos efforts pour adopter cette méthode n'ont pas abouti compte tenu du fait que nous ne disposions pas dans notre base de données des informations nécessaires pour modéliser le processus de sélection. En effet, pour mener à bien cette méthode, il nous fallait modéliser premièrement les deux processus de sélection, soit la probabilité qu'un prêteur soit ou non un multi-prêteur et la probabilité qu'une campagne réussisse ou échoue. Dans le premier cas, des informations sur les prêteurs (revenu, fonction, motivation…) sont primordiales. Malheureusement et comme c'est le cas sur la plupart des plateformes de financement participatif, très peu d'informations sont fournies sur les investisseurs. Dans le second cas, nous n'étions pas en mesure de satisfaire à la condition d'exclusion restrictive, sans laquelle la méthode de Heckman devient moins fiable que la méthode des moindres carrés ordinaires (Wolfolds et Siegel,2018).

Selon l'étude de Ly et Mason (2012b), la préférence des prêteurs pour les prêts de groupe varie en fonction du nombre des emprunteurs qui composent les groupes. Il serait intéressant, dans des recherches futures, d'examiner cela pour les multi-prêteurs. Prouver que les multi-prêteurs démontrent une certaine préférence pour les grands groupes par rapport à des petits groupes d'emprunteurs, permettrait de confirmer que l'objectif de ces prêteurs en choisissant des prêts de groupe est d'éteindre l'impact d'une seule contribution à plusieurs bénéficiaires.

Les résultats obtenus avec l'objectif 1 suggèrent qu'en fonction de leur niveau de diversification, le groupe des multi-prêteurs n'est pas parfaitement homogène. Une autre piste de recherche serait donc d'étudier le comportement d'investissement dans ces sous-groupes.

Bibliographie

Afsa, C. (2016). Le modèle Logit : Théorie et application. *Institut National de la Statistique et des Études Économiques*.

Algamal, Z. Y., & Lee, M. H. (2016). High dimensional logistic regression model using adjusted Elastic Net penalty. *ResearchGate*.

Ali, J., Khan, R., Ahmad, N., & Maqsood, I. (2012). Random Forests and Decision Trees. *International Journal of Computer Science Issues*.

Allison, T. H., McKenny, A. F., & Short, J. C. (2013). The effect of entrepreneurial rhetoric on microlending investment: An examination of the warm-glow effect. *Journal of Business 28*.

Allison, T., Short, J., & Webb, J. (2015). Crowdfunding in a prosocial microlending environment: examining the role of intrinsic versus extrinsic cues. *Entrepreneurship Theory and Practice 39*.

Anago, F. S., & Houngbeme, D. J.-L. (2015). Construction d'un indicateur synthétique de croissance inclusive. *Revue d'Analyse des politiques économiques et financières*, 69 - 105.

Anderberg, M. R. (1973). *Cluster Analysis for Applications.* New York: Academic Press.

Andreoni, J. (2007). Giving gifts to groups: How altruism depends on the number of recipients. *Journal of Public Economics*, 1731-1749.

Antonelli, J., & Cefalu, M. (2020). Averageing causal estimators in high dimensions. *Journal of Causal Inference*, 92-107.

Armendariz de Aghion, B. (1999). On the Design of a Credit Agreement with Peer Monitoring. *Journal of Development Economics*, 79-104.

Armendáriz de Aghion, B., & Morduch, J. (2005). *The Economics of Microfinance.* Massachuttes Institute of Technology.

Asselin, L.-M. (2009). Composite Indicator of Multidimensional Poverty.

Athey, S., Imbens, G., Pham, T., & Wager, S. (2017). Estimating Average Treatment Effects:Supplementary Analyses and Remaining Challenges. *The American Economic Review*.

Banerjee, A. V., Besley, T., & Guinnane, T. W. (1994). The Neighbor's Keeper: The Design of a Credit Cooperative with Theory and a Test. *Quarterly Journal of Economics*, 491–515.

Bashir, D., Montañez, G. D., Sehra, S., Segura, P. S., & Lauw, J. (2020). An Information-Theoretic Perspective on Overfitting and Underfitting.

Belleflamme , P., Lambert, T., & Schwienbacher, A. (2010). Crowdfunding: An Industrial Organization Perspective. *Digital Business Models: Understanding Strategies*, (pp. 1-30). Paris.

Belleflamme, P., Lambert, T., & Schwienbacher, A. (2012). Crowdfunding: Tapping the Right Crowd. *SSRN*.

Belloni, A., Chernozhukov, V., & Hansen, C. (2014). Inference on treatment effects after selection among high-dimensional controls. *The Review of Economic Studies*, 608–650.

Belloni, A., Chernozhukov, V., & Hansen, C. (2014). Inference on treatment effects after selection among high-dimensional controls. *The Review of Economic Studies*, 608-650.

Bernile, G., Bhagwat, V., & Yonker, S. (2018). Board diversity, firm risk, and corporate policies. *Journal of Financial Economics*, 588-612.

Berns, J. P., Figueroa-Armijos, M., da Motta Veiga, S. P., & Dunne, T. C. (2020). Dynamics of Lending-Based Prosocial Crowdfunding: Using a Social Responsibility Lens. *Journal of Business Ethics*, 169–185.

Besley, T., & Coate, S. (1995). Group Lending, Repayment Incentives and Social Collateral. *Journal of Development Economics*, 1-18.

Bierens, H. J. (2008). The Logit Model: Estimation, Testing and Interpretation.

Bligh, M. C., Kohles, J. C., & Meindl, J. R. (2004a). Charisma under crisis: Presidential leadership, rhetoric, and media responses before and after the September 11th terrorist attacks. *Leadership Quarterly*, 211-239.

Bligh, M. C., Kohles, J. C., & Meindl, J. R. (2004b). Charting the language of leadership: A methodological investigation of President Bush and the crisis of 9/11. *Journal of Applied Psychology*, 562-574.

Blossfeld, H., & Rohwer, G. (2002). *Techniques of Event History Modeling,New Approaches to Causal Analysis.* Mahwah, NJ : Lawrence Erlbaum Associates, ©2002.

Blume, M. E., Crockett, J., & Friend, I. (1974). Stockholders in the United States: Characteristics and Trends. *Survey of Current Business*, 16-40.

Bratton, M. (1986). Financing smallholder production: A comparison of individual and group credit schemes in Zimbabwe. *Public Administration and Development*, 115-132.

Breiman, L. (2001). *Random forests.* Statistics Department, University of California.

Breslow, N. (Biometrics). Covariance analysis of censored survival data. *1974*, 89-99.

Bretschneider, U., & Leimeister, J. M. (2017). Not just an ego-trip: Exploring backers' motivation for funding in incentive-based crowdfunding. *The Journal of Strategic Information Systems*.

Bruett, T. (2007). Cows, kiva, and prosper. com: How disintermediation and the internet are changing microfinance. *Community Development Investment Review*, 44-50.

Burtch, G., Ghose, A., & Wattal, S. (2013). An Empirical Examination of the Antecedents and Consequences of Contribution Patterns in Crowd-Funded Markets. *SSRN*.

Calvet, L. E., Campbell, J. Y., & Sodini, P. (2009). Measuring the Financial Sophistication of Households. *The American Economic Review*, 393–398.

Castrejon, L., Ballas, N., & Courville, A. (2019). Improved conditional VRNNs for video prediction. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)}.*

Chaffee, E. C., & Rapp, G. C. (2011). *Regulating On-Line Peer-to-Peer Lending in the Aftermath of Dodd-Frank: In Search of an Evolving Regulatory Regime for an Evolving Industry.* Washington and Lee Law Review.

Charrad, M., Ghazzali, N., Boiteau, V., & Niknafs, A. (2014). NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set. *Journal of Statistical Software*, 1-36.

Chatterji, A. K. (2009). Spawned with a silver spoon? Entrepreneurial performance and innovation in the medical device industry. *Strategic Management Journal*, 185- 206.

Chen, C., Liaw, A., & Breiman, L. (2004). Using Random Forest to Learn Imbalanced Data.

Cholakova, M., & Clarysse, B. (2014). Does the Possibility to Make Equity Investments in Crowdfunding Projects Crowd Out Reward-Based Investments? *Entrepreneurship Theory and Practice*.

Choy, K., & Schlagwein, D. (2016). Crowdsourcing for a Better World: On the Relation Between IT Affordances and Donor Motivations in Charitable Crowdfunding. *Information Technology & People*.

Christensen, M. (2018). A Key Rate of Sector and Country Diversification. *The Journal of Wealth Management*.

Collins, L., & Pierrakis, Y. (2012). *The venture crowd: Crowdfunding equity investment into business.* London: Nesta.

Cook, D. F., Ragsdale, C. T., & Major, R. L. (2000). Combining a neural network with a genetic algorithm for process parameter optimization. *Engineering Applications of Artificial Intelligence*, 391–396.

Cornelissen, J. P., & Clarke, J. S. (2010). Imagining and rationalizing opportunities: Inductive reasoning and the creation and justification of new ventures. *Academy of Management Review*, 539–557.

Cumming, D. J., & Johan, S. A. (2013). Demand Driven Securities Regulation: Evidence from Crowdfunding. *Venture Capital: An International Journal of Entrepreneurial Finance*, 361-379.

Dacin, P. A., & Dacin, T. M. (2010). Social entrepreneurship: why we don't need a new theory and how we move forward from here. *The Academy of Management Perspectives*, 37-57.

Davis, P. E. (1995). *Debt, Financial Fragility, and Systemic Risk.* Oxford: Clarendon Press.

Deci, E. L. (1997). *Intrinsic Motivation.* New York: Plenum.

DeYoung, R., Sprong, K., & Sullivan, R. J. (2001). Who's minding the store? Motivating and monitoring hired managers at small closely held commercial banks. *Journal of Banking and Finance*, 1209–1244.

Diamond, D. W. (1984). Financial intermediation and delegated monitoring. *Review of Economic Studies*, 393–414.

Dimmock, S. G., Kouwenberg, R., Mitchell, O. S., & Peijnenburg, K. (2020). Household Portfolio Underdiversification and Probability Weighting: Evidence from the Field. *The Review of Financial Studies*.

Ding, P., & Li, F. (2018). Causal Inference: A Missing Data Perspective. *Statistical Science*, 214–237.

Dougherty, C. (2007). *Introduction to Econometrics.* New York: Oxford University Press.

Duvendack, M., Palmer-Jones, R., Copestake, J. G., Hooper, L., Loke, Y., & Rao, N. (2011). *What is the evidence of the impact of microfinance on the well-being of poor people?* London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.

Efron, B. (1977). The efficiency of Cox's likelihood function for censored data. *Journal of the American Statistical Association*, 557-565.

Elder, J. (2012). *Multilayer perceptrons.* University of Michigan Electrical & Computer Engineering Division.

Emiris, M. (2004). Sectoral vs. country diversification benefits and downside risk. *National Bank of Belgium*.

Ertefaie, A., Asgharian, M., & Stephens, D. A. (2018). Variable selection in causal inference using a simultaneous penalization method. *Journal of Causal Inference*.

Estrin, S., Gozman, D., & Khavul, S. (2018). The evolution and adoption of equity crowdfunding: entrepreneur and investor entry into a new market. *Small Business Economics*, 425–439.

Ettaouil, M., & Ghanou, Y. (2009). Neural architectures optimization and genetic algorithms.

Eysa, S., & Saeed, G. (2005). Optimum design of structures by an improved genetic algorithm using neural networks. *Advances in Engineering Software*, 757–767.

Flanagin, A. J. (2007). Commercial markets as communication markets: Uncertainty reduction through mediated information exchange in online auctions. *New Media & Society*, 401-423.

Freedman, S. M., & Jin, G. Z. (2011). Learning by Doing with Asymmetric Information: Evidence from Prosper.com. *National Bureau of Economic Research*.

Galak, J., Small, D. A., & Stephen, A. T. (2011). Micro-Finance Decision Making: A Field Study of Prosocial Lending. *SSRN*.

Gallagher, M., Moustafa, N., & Lakshika, E. (2020). Advances in Artificial Intelligence. *Lecture Notes in Computer Science.*

Gao, Q., & Lin, M. (2015). Lemon or Cherry? The Value of Texts in Debt Crowdfunding. *Working Paper*.

Gentzkow, M., Shapiro, J. M., & Taddy, M. (2015). Measuring polarization in high dimensional data: Method and application to congressional speech. *Technical Report*.

Gerber, E., & Hui, J. (2014). Crowdfunding: Motivations and Deterrents for Participation. *ACM Transactions on Computer-Human Interaction*.

Ghatak, M. (1999). Group lending, local information and peer selection. *Journal of Development Economics*, 27–50.

Ghatak, M. (2000). Screening by the company you keep: Joint liability lending and the peer selection effect. *The Economic Journal*, 601-631.

Ghatak, M., & Guinnane, T. W. (1999). The Economics of Lending with Joint Liability: Theory and Practice. *Economic Growth Center, Yale University*.

Glaeser, E. L., & Shleifer, A. (2001). Not-for-Profit Entrepreneurs. *Journal of Public Economics*.

Goetzmann, W. N., & Kumar, A. (2008). Equity Portfolio Diversification. *Review of Finance*.

Goetzmann, W. N., Li, L., & Rouwenhorst, K. G. (2005). Long-term global market correlations. *Journal of Business*, 1-38.

Graves, A. (2012). Long Short-Term Memory. *Supervised Sequence Labelling with Recurrent Neural Networks*, 37–45.

Greene, W. H. (2008). *Econometric Analysis.* New Jersey: Pearson Prentice Hall.

Guo, S. Y., & Fraser, M. W. (2014). *Propensity Score Analysis: Statistical Methods and Applications.* Sage Publications, Inc.

Hahn, P. R., Carvalho, C., & Puelz, D. (2016). Bayesian regularized regression for treatment effect estimation from observational data. *SSRN*.

Hahn, P. R., Carvalho, C., & Puelz, D. (2016). Bayesian regularized regression for treatment effect estimation from observational data. *SSRN*.

Haider, A. (2014). Methods to avoid over-fitting and under-fitting in supervised machine learning (comparative study).

Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (1992). *Multivariate data analysis with readings.* Englewood Cliffs: Prentice Hall.

Harpham, C., Dawson, C. W., & Brown, M. R. (2004). A review of genetic algorithms applied to training radial basis function networks. *Neural Computing & Applications*, 193-201.

Harsh, B., & Surbhi, B. (2012). Application of Genetic Algorithms in Machine learning.

Hart, R. P. (1984). *Verbal Style and the Presidency.* New York: Academic Press.

Hart, R. P. (2002). *Campaign Talk: Why Elections Are Good for Us.* New York: Princeton University Press.

Hart, R. P. (2010). *DICTION 6.0 Manual.* Austin, Texas: Digitext, Inc.

Hart, R. P., & Jarvis, S. E. (1997). Political debate: Forms, styles and media. *American Behavioral Sciences*.

Hassan, R., Mohammed, A., Idrissi, J., Amine, M., Youssef, G., & Mohamed, E. (2016). Multilayer Perceptron: Architecture Optimization and Training. *International Journal of Interactive Multimedia and Artificial Intelligence*.

Hastie, T., Tibshirani, R., & Friedman, J. (2017). *The Elements of Statistical Learning.* Springer.

Haugen, R. A. (2001). *Modern Investment Theory.* New Jersey: Prentice Hall.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica: Journal of the econometric society*, 153-161.

Herzenstein, M., Andrews, R. L., Dholakia, U. M., & Lyandres, E. (2008). The Democratization Of Personal Consumer Loans? Determinants Of Success In Online Peer-To-Peer Lending Communities.

Herzenstein, M., Sonenshein, S., & Dholakia, U. M. (2011). Tell Me a Good Story and I May Lend You My Money: The Role of Narratives in Peer-to-Peer Lending Decisions. *SSRN*.

Hinton, G. E., Srivastava, N., Sutskever, I., & Salakhutdinov, R. R. (2012). Improving neural networks by preventing co-adaptation of feature detectors. *arXiv*.

Ho, T. K. (1998). The random subspace method for constructing decision forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 832-844.

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, 1735–1780.

Howe, J. (2006). The Rise of Crowdsourcing. *Scientific Research Publishing*.

Islam, M. M., & Yao, X. (2008). Evolving artificial neural network ensembles. *IEEE Computational Intelligence Magazine*.

Jadid, S., Said, J. A., Hitham, A., Muhammad, N., & Asim, E. (2018). Long Short Term Memory Recurrent Network for Standard and Poor's 500 Index Modelling. *International Journal of Engineering and Technology*, 25-29.

Jaume, R. H. (2017). The collaborative economy for channelling institutional and household savings. *The Spanish Review of Financial Economics*, 12-20.

Jean, N., Burke, M., Xie, M., Davis, W. M., Lobell, D. B., & Ermon, S. (2016). Combining satellite imagery and machine learning to predict poverty. *Science*, 790-794.

Joo-Kitano, H. (2021). The Role of Expertise and Risk Diversification Strategies in Equity Crowdfunding. *The Journal of International DBA Studies*.

Kang, J. S., Kuznetsova, P., Luca, M., & Choi, Y. (2013). Where not to eat? improving public policy by predicting hygiene inspections using online reviews. *Conference on Empirical Methods in Natural Language Processing*, (pp. 1443-1448).

Keith, K., Jensen, D., & O'Connor, B. (2020). Text and causal inference: A review of using text to remove confounding from causal estimates. *ACL*.

Kennedy, P. (2008). *A Guide to Econometrics.* Malden: Blackwell Publishing.

Ketchen, D. J., & Shook, C. L. (1996). The Application of Cluster Analysis in Strategic Management Research: An Analysis and Critique. *Strategic Management Journal*, 441–458.

Khavul, S. (2010). Microfinance: Creating Opportunities for the Poor? *Academy of Management Perspectives*.

Kim, P., Buffart, M., & Croidieu, G. (2016). TMI: Signaling Credible Claims in Crowdfunding Campaign Narratives. *Group & Organization Management*, 717-750.

Kirby, E., & Worner, S. (2014). *Crowd-funding: An Infant Industry Growing Fast.*

Klafft, M. (2008a). Online Peer-to-Peer Lending: A Lenders' Perspective. *SSRN*.

Knill, A. (2009). Should Venture Capitalists Put All Their Eggs in One Basket? Diversification versus Pure-Play Strategies in Venture Capital. *Financial Management*.

Kogut, T., & Ritov, I. (2005a). The 'Identified Victim' Effect: An Identified Group or Just an Individual? *Journal of Behavioral Decision Making*.

Kogut, T., & Ritov, I. (2005b). The Singularity Effect of Identified Victims in Separate & Joint Evaluation. *Organizational Behavior and Human Decision Processes*.

Kolluri, J., Kotte, V. K., Phridviraj, M., & Razia, S. (2020). Reducing Overfitting Problem in Machine Learning Using Novel L1/4 Regularization Method. *International Conference on Trends in Electronics and Informatics*, 934-938.

Kuang, K., Li, L., Geng, Z., Xu, L., Zhang, K., Liao, B., . . . Jiang, Z. (2020). Causal Inference. *Engineering*, 253-263.

Lambert, T., & Schwienbacher, A. (2010). An Empirical Analysis of Crowdfunding. *Louvain School of Management*.

Landis, J. R., & Koch, G. G. (1977). An Application of Hierarchical Kappa-Type Statistics in the Assessment of Majority Agreement among Multiple Observers. *Biometrics*, 363-374.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*.

Larrimore, L., Jiang, L., Larrimore, J., Markowitz, D., & Gorski, S. (2011). Peer to Peer Lending: The Relationship Between Language Features, Trustworthiness, and Persuasion Success. *Journal of Applied Communication Research*, 19-37.

Le, Q. V. (2015). *Nonlinear Classifiers and The Backpropagation Algorithm.*

Leslie, S. (2018). A disciplined approach to neural network hyper-parameters: Part 1 -- learning rate, batch size, momentum, and weight decay. *arXiv*.

Levy, O., Goldberg, Y., & Dagan, I. (2015). Improving distributional similarity with lessons learned from word embeddings. *Transactions of the Association for Computational Linguistics*, 211– 225.

Li, L. (2011). Propensity Score Analysis with Matching Weights. *arXiv*.

Liaw, A., & Wiener, M. (2002). Classification and Regression by randomForest. *ResearchGATE*.

Lin, Y., & Boh, W. (2019). How Different Are Crowdfunders? Examining Archetypes of Crowdfunders. *Journal of the Association for Information Science and Technology*.

Lin, Y., Boh, W., & Goh, K. (2014). How Different are Crowdfunders? Examining Archetypes of Crowdfunders and Their Choice of Projects. *SSRN Electronic Journal*.

Lins, A., & Ludermir, T. (2013). *lins-optimization.*

Liu, Y., Chen, R., Chen, Y., Mei, Q., & Salib, S. (2012). I loan because...: Understanding motivations for pro-social lending. *Proceedings of the fifth ACM international conference on Web search and data mining*, 503-512.

Loye, G. (2019). *Attention Mechanism.*

Ludermir, T. B. (2005). Hybrid Optimization Algorithm for the Definition of MLP Neural Network Architectures and Weights. *Proceedings of the Fifth International Conference on Hybrid Intelligent Systems*.

Lukkarinen, A., Teich, J. E., & Wallenius, H. (2016). Success drivers of online equity crowdfunding campaigns. *Decision Support Systems*.

Lunceford, J. K., & Davidian, M. (2004). Stratification and weighting via the propensity score in estimation of causal treatment effects: a comparative study. *Statistics in medicine*.

Luong, M.-T., Pham, H., & Manning, C. D. (2015). Effective Approaches to Attention-based Neural Machine Translation. *Empirical Methods in Natural Language Processing.*

Ly, P., & Mason, G. (2012b). Competition Between Microfinance NGOs: Evidence from Kiva. *World Development*, 643–655.

Ly, P., & Mason, G. (2012b). Individual preferences over development projects: Evidence from microlending on kiva. *International Journal of Voluntary and Nonprofit Organizations*, 1036–1055.

Makino, S., Isobe, T., & Chan, C. M. (2004). Does country matter? *Strategic Management Journal*, 1027-1043.

Manashty, A., & Thomson, J. L. (2017). A New Temporal Abstraction for Health Diagnosis Prediction using Deep Recurrent Networks. *Proceedings of the 21st International Database Engineering & Applications Symposium* (pp. 14-19). New York: Association for Computing Machinery.

Markozannes, G., Vourli, G., & Ntzani, E. (2021). A survey of methodologies on causal inference methods in meta-analyses of randomized controlled trials. *Systematic Reviews*.

Martens, M. L., Jennings, J., & Jennings, P. D. (2007). Do the stories they tell get them the money they need? The role of entrepreneurial narratives in resource acquisition. *Academy of Management Journal*, 1107–1132.

McKinnon, S. L., Dickinson, E., Carr, J. N., & Chávez, K. R. (2013). Kiva.org, Person-to-Person Lending, and the Conditions of Intercultural Contact. *Howard Journal of Communications*.

Mélot, C. (2008). Logistic Regression versus Cox Regression.

Merton, R. C. (1969). Lifetime Portfolio Selection Under Uncertainty: The Continuous Time Case. *Review of Economics and Statistics*, 247-257.

Milligan, G. W. (1980). An examination of the effect of six types of error perturbation on fifteen clustering algorithms. *Psychometrika*, 325-342.

Mintjes, D. B. (2016). The role of financial literacy on portfolio diversification in peer-to-peer lending. *Business Economics*.

Mollick, E. (2014). The Dynamics of Crowdfunding: An Exploratory Study. *Journal of Business Venturing*, 1-16.

Moonhee, C., & Gawon, K. (2017). A cross-cultural comparative analysis of crowdfunding projects in the United States and South Korea. *Computers in Human Behavior*, 312-320.

Morris, M., Webb, J., & Franklin, R. (2011). Understanding the Manifestation of Entrepreneurial Orientation in the Nonprofit Context. *Entrepreneurship Theory & Practice*.

Moss, T., Neubaum, D., & Meyskens, M. (2015). The Effect of Virtuous and Entrepreneurial Orientations on Microfinance Lending and Repayment: A Signaling Theory Perspective. *Entrepreneurship Theory and Practice*.

Moss, T., Renko, M., Block, E., & Meyskens, M. (2017). Funding the story of hybrid ventures: Crowdfunder lending preferences and linguistic hybridity. *Journal of Business Venturing*.

Moysidou, K. (2016). Motivations to contribute financially to crowdfunding projects. *World Scientific*.

Nagy, I. (2026). Crowdlending: un placement adapté pour diversifier son épargne ? *Haute École de Gestion Genève*.

Nardo, M., Saisana, M., Saltellii, A., Tarantola, S., Hoffman, A., & Giovannini, E. (2005). *Handbook on Constructing Composite Indicators: Methodology and User Guide.* Paris: OECD Statistics Working Papers.

Norton, E., & Tenenbaum, B. H. (1993). Specialization versus diversification as a venture capital investment strategy. *Journal of Business Venturing*, 431-442.

Ordanini, A., Miceli, L., Pizzetti, M., & Parasuraman, A. (2011). Crowdfunding: Transforming customers into investors through innovative service platforms. *Journal of Service Management*, 443–470.

Parhankangas, A., & Renko, M. (2017). Linguistic style and crowdfunding success among social and commercial entrepreneurs. *Journal of Business Venturing*, 215-236.

Paxton, J. (1996). Determinants of Successful Group Loan Repayment: An Application to Burkina Faso. *Doctoral Dissertation,Ohio State University*.

Paxton, J., Graham, D., & Thraen, C. (2000). Modeling group loan repayment behavior: new insights from Burkina Faso. *Economic Development & Cultural Change*, 639–655.

Payne, G. T., Brigham, K. H., Broberg, J. C., Moss, T. W., & Short, J. C. (2011). Organizational virtue orientation and family firms. *Business Ethics Quarterly*, 257–285.

Pennington, J., Socher, R., & Manning, C. D. (2014). Glove: Global vectors for word representation. *Empirical Methods in Natural Language Processing*, 1532–1543.

Pereyra, G., Tucker, G., Chorowski, J., Kaiser, Ł., & Hinton, G. (2017). Regularizing neural networks by penalizing confident output distributions. *ICLR workshops*.

Perez, L., & Wang, J. (2017). The effectiveness of data augmentation in image classification using deep learning. *arXiv*.

Peterson, M. F., Arregle, J. L., & Martin, X. (2012). Multilevel models in international business research. *Journal of International Business Studies*, 451-457.

Priberny, C., & Sommerc, W. (2021). What drives the loan funding on the peer-to-peer platform Kiva? *World Finance Conference 2021.*

Punj, G., & Stewart, D. W. (1983). Cluster Analysis in Marketing Research: Review and Suggestions for Application. *Journal of Marketing Research*.

Renko, M. (2013). Early challenges of nascent social entrepreneurs. *Entrepreneurship Theory and Practice*, 1045–1069.

Robins, J. M., & Rotnitzky, A. (1995). Semiparametric Efficiency in Multivariate Regression Models with Missing Data. *Journal of the American Statistical Association*, 122–129.

Robins, J. M., Rotnitzky, A., & Zhao, L. P. (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association*.

Rodriguez, M. Z., Comin, C. H., Casanova, D., Bruno, O. M., Amancio, D. R., Costa, L. F., & Rodrigues, F. A. (2019). Clustering algorithms: A comparative approach. *PLoS One*.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 41-55.

Rosenblatt, F. (1958). The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain. *Psychological Review*, 386-408.

Rossi, S. P., Schwaiger, M. S., & Winkler, G. (2009). How loan portfolio diversification affects risk, efficiency and capitalization: A managerial behavior model for Austrian banks. *Journal of Banking & Finance*, 2218-2226.

Rubin, D. (688–701). Estimating causal effects of treatments in randomized and non-randomized studies. *Journal of Educational Psychology*, 1974.

Rubin, D. B. (2005). Causal inference using potential outcomes: Design, modeling, decisions. *Journal of the American Statistical Association*, 322–331.

Rulli, E., Ghilotti, F., Biagioli, E., Porcu, L., Marabese, M., D'Incalci, M., . . . Torri, V. (2018). Assessment of proportional hazard assumption in aggregate data: a systematic review on statistical methodology in clinical trials using time-to-event endpoint. *British Journal of Cancer*, 1456–1463.

Samuelson, P. A. (1969). Lifetime Portfolio Selection by Dynamic Stochastic Programming. *Review of Economics and Statistics*.

Scharfstein, D. O., Rotnitzky, A., & Robins, J. M. (1999). Adjusting for nonignorable drop-out using semiparametric nonresponse models (with discussion and rejoinder). *Journal of the American Statistical Association*, 1096–146.

Schemper, M. (1992). Cox analysis of survival data with non-proportional hazard functions. *The Statistician*, 455–465.

Schwienbacher, A., & Larralde, B. (2010). Crowdfunding of small entrepreneurial ventures. *Social Science Research Network*.

Sharpe, W. F. (1964). Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk. *Journal of Finance*, 425-442.

Shepherd, B. (2010). Geographical Diversification of Developing Country Exports. *World Development*, 1217-1228.

Short, J. C., Broberg, J. C., Cogliser, C. C., & Brigham, K. H. (2010). Construct validation using computer-aided text analysis (CATA). *Organization Research Methods*, 320–347.

Short, J. C., Moss, T. W., & Lumpkin, G. T. (2009). Research in social entrepreneurship: Past contributions and future opportunities. *Strategic Entrepreneurship Journal*, 161–194.

Short, J. C., Payne, G. T., Brigham, K. H., Lumpkin, G. T., & Broberg, J. C. (2009). Family businesses and entrepreneurial orientation in publicly-traded businesses: A comparative analysis of the S&P 500. *Family Business Review*, 9-24.

Shortreed, S. M., & Ertefaie, A. (2017). Outcome-adaptive lasso: Variable selection for causal inference. *Biometrics*.

Shortreed, S. M., & Ertefaie, A. (2017). Outcome-adaptive lasso: Variable selection for causal inference. *Biometrics*, 1111-1122.

Sirignano, J. A., Sadhwani, A., & Giesecke, K. (2016). Deep Learning for Mortgage Risk. *arXiv*.

Small, D. A., & Loewenstein, G. (2003). Helping a Victim or Helping the Victim: Altruism and Identifiability. *Journal of Risk and Uncertainty*.

Sridhar, D., & Getoor, L. (2020). Estimating causal effects of tone in online debates. *International Joint Conference on Artificial Intelligence*.

Sridhar, D., Springer, A., Hollis, V., Whittaker, S., & Getoor, L. (2018). Estimating causal effects of exercise from mood logging data. *IJCAI/ICML Workshop on CausalML*.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*.

Stensrud, M. J., & Hernán, M. A. (2020). Why test for proportional hazards? *Journal of the American Medical Association*, 1401–1402.

Stiglitz, J. E. (1990). Peer monitoring and credit markets. *World Bank Economic Review*, 351–366.

Stiroh, K. (2004). Diversification in banking: Is non-interest income the answer? *Journal of Money, Credit and Banking*, 853–882.

Stiroh, K., & Rumble, A. (2006). The dark side of diversification: The case of US financial holding companies. *Journal of Banking and Finance*, 2131–2161.

Storey, D. J. (2000). *Small business critical perspectives on business and management.* New York: Routledge.

Sun, Y. (2021). Optimization and Simulation of Manuscript Management System Based on Fuzzy Genetic Neural Network. *Computational Intelligence and Neuroscience*.

Swanson, A. (2016). Geographic Diversification and Commercial Bank Insolvency Risk and Essays on Crowdfunding and Subscription Games. *Conference Proceedings*.

Tausczik, Y. R., & Pennebaker, J. W. (2010). The Psychological Meaning of Words LIWC and Computerized Text Analysis Methods. *Journal of Language and Social Psychology*, 24-54.

Therneau, T. M., & Grambsch, P. M. (2000). *Modeling Survival Data: Extending the Cox Model.* New York: Springer.

Tian, H., Chen, S., & Shyu, M. (2019). Genetic Algorithm Based Deep Learning Model Selection for Visual Data Classification. *IEEE 20th International Conference on Information Reuse and Integration for Data Science*, (pp. 127-134).

Tibshirani, R. (2006). *Regularization : Ridge Regression and the LASSO.*

Vallerand, R. J. (1997). Toward A Hierarchical Model of Intrinsic and Extrinsic Motivation. *Advances in experimental social psychology*.

Van Der Laan, M. J., & Rubin, D. (2006). Targeted maximum likelihood learning. *The International Journal of Biostatistics*.

Van Horne, J. C., Blume, M. E., & Friend, I. (1975). The asset structure of individual portfolios and some implications for utility functions. *Journal of Finance*, 585–603.

Veitch, V., Sridhar, D., & Blei, D. M. (2020). Adapting text embeddings for causal inference. *UAI*.

Venkatesan, D., Kannan, K., & Saravanan, R. (2009). A genetic algorithm-based artificial neural network model for the optimization of machining processes. *Neural Computing & Applications*, 135–140.

Venkatesh, V. (1999). Creation of Favorable User Perceptions: Exploring the Role of Intrinsic Motivation. *Mechanical Engineering eJournal*.

Wang, H., Greiner, M., & Aronson, J. E. (2009). People-to-People Lending: The Emerging E-Commerce Transformation of a Financial Market. *SpringerLink*.

Weld, G., West, P., Glenski, M., Arbour, D., Rossi, R., & Althoff, T. (2020). Adjusting for confounders with text: Challenges and an empirical evaluation framework for causal inference. *arXiv*.

Wenzlaff, K., Gumpelmaier, W., & Eisfeld-Reschke, J. (2012). Definition von crowdfunding. *Institut für Kommunikation in sozialen Medien*.

Wijayatunga, P. (2014). Causal Effect Estimation Methods. *Journal of Statistical and Econometric Methods*, 153-170.

Wilson, A., & Reich, B. J. (2014). Confounder selection via penalized credible regions. *Biometrics*, 852–861.

Wolfolds, S., & Siegel, J. (2018). Misaccounting for endogeneity: the Peril of relying on the Heckman two-step method without a valid Instrument. *Strategic Management Journal*.

Wooldridge, J. M. (2009). *Introductory Econometrics: A Models Approach.* New York: South-Western Publishers.

Ying, X. (2019). An overview of overfitting and its solutions. *Journal of Physics*.

Yum, H., Lee, B., & Chae, M. (2012). From the wisdom of crowds to my own judgment in microfinance through online peer-to-peer lending platforms. *Commer*.

Yunus, M. (2007). *Creating a World without poverty: Social business and the future of capitalism.* New York: Public Affairs.

Yuthas, K., Rogers, R., & Dillard, J. F. (2002). Communicative action and corporate annual reports. *Journal of Business Ethics*, 141-157.

Zachary, R. P., McKenny, A., Short, J. C., & Payne, G. T. (2011). Family Business and Market Orientation Construct Validation and Comparative Analysis. *Family Business Review*, 233-251.

Zaremba, W., Sutskever, I., & Vinyals, O. (2015). Recurrent neural network regularization.

Zeller, M. (1998). Determinants of Repayment Performance in Credit Groups: The Role of Program Design, Intragroup Risk Pooling, and Social Cohesion. *Economic Development and Cultural Change*, 599–620.

Zhang, H., Zhang, L., & Jiang, Y. (2019). Overfitting and Underfitting Analysis for Deep Learning Based End-to-end Communication Systems. *11th International Conference on Wireless Communications and Signal Processing (WCSP)*, (pp. 1-6).

Zhang, J., & Liu, P. (2012). Rational Herding in Microloan Markets. *Management Science*, 892-912.

ZINSOU, T. T. (2008). Construction d'un indicateur synthétique de conjoncture industrielle. *Memoire online*.

Zott, C., & Huy, Q. N. (2007). How entrepreneurs use symbolic management to acquire resources. *Administrative Science Quarterly*, 70–105.

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, 1418–1429.

Zou, H., & Hastie, T. (2004). Regularization and variable selection via the elastic net. *J. R. Statist. Soc. B*.

Zouitene, H. (2020). Inférence causale pour mesurer le retour sur les investissements publicitaires. Maîtrise en statistique - avec mémoire. Retrieved from CorpusUL.

Annexe A : Description des variables

Variables	Description	Dictionnaire de mots
Variables dépendantes		
Duree	Nombre de jours séparant la date de publication de la demande prêt sur le site de Kiva et la date à laquelle les fonds ont été levés	---
Succes	Variable binaire qui prend la valeur 1 si le projet a été totalement financé et 0 sinon	---
Variables indépendantes		
Classe	Variable catégorielle à trois modalités faible, moyenne et élevé	---
Profit	Score mesurant le langage de profit utilisé dans les récits de demande de prêts (une valeur élevée pour le score indique un niveau élevé du langage dans le récit)	Beneficial, benefit, benefited, benefits, cash, cost effective, cost effectiveness, cost efficient, desirable, desire, desired, earn, earning, earnings, earns, emolument, fecundity, fructuous, fruit, fruitful, fruitfully, fruits, gain, gained, gainful, gaining, gains, generate, generates, generating, generative, income, incomes, lucrative, lucre, money, moneymaking, net income, proceeds, productive, productivity, profit, profit, making, profitable, profits, profited, profiting, propitious, prosper, prospered, prospering, prosperous, prospers, return, revenue, reward, rewarded, rewarding, rewards, rich, valuable, value, win, winnings, wins, yield, yielding, yields, paid off, pay off, paid dividends, pay dividends, revenues, bottom line, EBIT, EBITDA, income
Risque	Score mesurant le langage de risque utilisé dans les récits de demande de prêts (une valeur élevée pour le score indique un niveau élevé du langage dans le récit)	adventuresome, adventurous, audacious, bet, bold, bold-spirited, brash, brave, chance, chancy, courageous, danger, dangerous, dare, daredevil, daring, dauntless, dicey, enterprising, fearless, gamble, gutsy, headlong, incautious, intrepid, plunge, precarious, rash, reckless, risk, risky, stake, temerity, uncertain, venture, venturesome, wager
Interet_humain	Score mesurant le langage d'intérêt humain utilisé dans les récits de demande de prêts (une valeur élevée pour le score indique un niveau élevé du langage dans le récit)	Pronoms personnels (he, his, ourselves, them), Membres de la famille et parents (cousin, wife, grandchild, uncle), termes génériques (friend, baby, human, persons).
Diversite	Score mesurant le niveau de diversité du langage utilisé dans les récits de demande de prêts (une valeur élevée pour le score indique un niveau élevé du langage dans le récit)	Mots décrivant des individus ou des groupes d'individus qui diffèrent de la norme. Ce caractère distinctif peut être comparativement neutre (inconsistent, contrasting, non-conformist) mais il peut aussi être positif ((exceptional, unique, individualistic) et négatif (illegitimate, rabble-rouser, extremist). D'un point de vue fonctionnel, l'hétérogénéité peut être un atout ((far-flung, dispersed, diffuse) ou un handicap (factionalism, deviancy, quirky), tout comme ses

Variables	Description	Dictionnaire de mots
		caractérisations : rare vs. queer, variety vs. jumble, distinctive vs. disobedient.
Preteur	Variable binaire qui prend la valeur de 1 si un multi-prêteur a participé au financement du projet et 0 sinon	---
Groupe	Variable binaire qui prend la valeur 1 si l'emprunteur est un groupe et 0 sinon	
Variables de contrôle		
Pays	Variable catégorielle désignant les pays des emprunteurs	---
Secteur	Variable catégorielle désignant les secteurs d'activité	---
Montant_pret	Montant de prêt en $ US	---
Duree_pret	Nombre de mois prévus au bout duquel le prêt doit être remboursé	---
Risque_IMF	Niveau de risque du partenaire de terrain mesuré par les cotes de risque variant entre 0 et 4. 0 indique un faible niveau de risque et 4 un niveau de risque élevé.	---
Couverture	Variable à 3 catégories (0, 1 et 2) indiquant comment le risque de taux de change est couvert	---
Tenacite	Score mesurant le niveau de ténacité dans les récits de demande de prêts (une valeur élevée pour le score indique un niveau élevé du langage dans le récit)	Tous les emplois du verbe être (is, am, will, shall) trois formes verbales définitives (has, must, do) et leurs variantes, ainsi que toutes les contractions associées (he'll, they've, ain't). Ces verbes évoquent la confiance et la totalité.
Nivellement	Score mesurant l'importance accordée par l'emprunteur aux personnes qui présentent des particularités (une valeur élevée pour le score indique un niveau élevé du langage dans le récit)	Mots utilisés pour ignorer les différences individuelles et pour donner un sentiment de complétude et d'assurance. On y trouve des termes totalisants (everybody, anyone, each, fully), des adverbes de permanence (always, completely, inevitably, consistently) et des adjectifs résolus (unconditional, consummate, absolute, open-and-shut).
Blame	Score mesurant à quel point l'emprunteur assume ses erreurs (une valeur élevée pour le score indique un niveau élevé du langage dans le récit)	Les termes désignant l'inconvenance sociale (mean, naive, sloppy, stupid) ainsi que le mal absolu (fascist, blood-thirsty, repugnant, malicious) composent ce dictionnaire. On y trouve également des adjectifs décrivant des circonstances malheureuses (bankrupt, rash, morbid, embarrassing) ou des vicissitudes imprévues (weary, nervous, painful, detrimental). Le dictionnaire contient également des dénigrements purs et simples : cruel, illegitimate, offensive, miserly.
Accomplissement	Score mesurant l'importance accordée aux	Mots exprimant la réalisation d'une tâche (establish, finish, influence, proceed) et le

Variables	Description	Dictionnaire de mots
	réalisations dans les récits (une valeur élevée pour le score indique un niveau élevé du langage dans le récit)	comportement humain organisé (buy, produce, employees, sell). Comprend les termes capitalistes (grow, increase, generate, construction), les modes d'expansion (handling, strengthen, succeed, outputs) et la fonctionnalité générale (gérer, renforcer, réussir, résultats). Comprend également le langage programmatique : agenda, enacted, working, leadership.
Preo_actuelle	Score mesurant la mention des préoccupations actuelles dans les récits de demande de prêts (une valeur élevée pour le score indique un niveau élevé du langage dans le récit)	Une liste sélective de verbes au présent extrapolée à partir de la liste de termes généraux et imagés de C. K. Ogden, qui apparaissent tous très fréquemment dans l'anglais américain standard. Le dictionnaire n'est pas axé sur un sujet particulier, mais renvoie plutôt à l'activité physique générale (cough, taste, sing, take), aux opérations sociales (canvass, touch, govern, meet) et à l'exécution d'une tâche (make, cook, print, paint).
Caractere_concret	Score mesurant le caractère concret des récits de demande de prêts (une valeur élevée pour le score indique un niveau élevé du langage dans le récit)	Un grand dictionnaire ne possédant aucune unité thématique autre que la tangibilité et la matérialité. On y trouve des unités sociologiques (farmers, African-Americans, Catholics), des groupes professionnels (carpenter, manufacturer, policewoman) et des alignements politiques (communistes, congressistes, européens). On y trouve également des structures physiques (palais de justice, temple, magasin), des formes de divertissement (television, soccer, CD-ROM), des termes de comptabilité (mortgage, salaries, finances) et des modes de transport (plane, boat, bicycle). En outre, le dictionnaire inclut les parties du corps (stomach, eyes, lips), les articles d'habillement (pants, shirt), les animaux domestiques (cat, insects, horse) et les denrées alimentaires (wine, cereals, sugar), et les éléments généraux de la nature (oil, silk, sand).
Variete	Score mesurant le niveau de variété linguistique utilisé dans les récits de demande de prêts (une valeur élevée pour le score indique un niveau élevé du langage dans le récit)	Cette mesure est conforme au ratio Type-Token de Wendell Johnson (1946) qui divise le nombre de mots différents dans un passage par le nombre total de mots du passage. Un score élevé indique que le locuteur évite l'exagération et préfère les déclarations précises et moléculaires.
Risque_defaut	Variable binaire qui prend la valeur 1 si le risque de défaut est supporté par le prêteur et 0 sinon	---
Récit	Variable texte fourni par les emprunteurs pour se présenter et présenter leur projet	---
Genre	Variable binaire qui prend la valeur 1 si l'emprunteur est une femme et 0 si l'emprunteur est un homme	---

Annexe B : Exemples de récits

Variables	Niveau	Score	Exemple
Profit	Min	0	R. L. has a gardening store located on the road at the entrance of Catarina. He sells various types of plants. He is single and thanks to his business he is able to support himself and help his family. He is requesting a loan from CEPRODEL to purchase various planters and garden plants because he already has various orders from customers.
	Max	59,52	José has decided to take out another loan in order to pay for labourers to bring in the harvest and to transport it to his house. This will save José time and so he will earn **money** quicker. He can then pay household expenses and **pay off** his debt. With his previous loan José has enough **money** to increase his harvest and to have disease-free crops. José increased his **income**, which he used to improve his family's living conditions and to pay off his **debt**.
Risque	Min	0	Koffi is 22 years old, single and has no children. Koffi has been a small-scale retailer since 2010. He buys telephone recharge cards and retails them to his varied clientele. He buys from communications companies in Lomé. Koffi is asking WAGES for his first loan to buy various sorts of telephone cards.
	Max	18,99	Rosalba, 61, is an enterprising woman who makes clothing. She has a beautiful family and four children. Rosalba lives in the town of Urrao, a **post-conflict** region very affected by the **violence** and high **poverty** rates, which is why the loan is higher **risk**. Rosalba wants to request a loan to buy a serger, fabric, zippers, buttons, and thread to expand her production, improve sales, and have the chance to get ahead and improve her beloved family's living conditions.
Interet_humain	Min	4,35	Hello, world! Meet Moses, an aspiring 20-year-old farmer from Motosiet village. At a tender age, Moses has embraced the spirit of entrepreneurship and is in farming to feed the rapidly growing population of Kenya. Moses' main income source is milk. This year, he is exploring other opportunities in the agricultural sector. Moses is seeking a KES 30,000 loan from Kiva lenders to venture into poultry farming. Poultry farming has so many products and byproducts to offer. Poultry industry growth is also driven by the expanding hotel and tourism industry. Moses believes there is so much to expect from this business move. Profits from this investment will be rolled back into the business to fuel growth.
	Max	58,82	This is Azamjon. He was born in 1990. He is a very kind and responsible man. He is married and has two **children**. His **wife** deals with animal husbandry. Azamjon doesn't work anywhere and deals with animal husbandry. He lives with his **parents**. Her parents are very kind. They love Azamjon. They live with a salary and live poorly. He loves his parents very much. He is the only son in his family. He takes care of them. His father is ill. He wants to get a loan and treat his **father**. He must buy medicine for his father's treatment. He needs money. He is asking for a loan 3,000 somoni with this aim.He awaits your support and mercy.
Diversite	Min	0	Ubania del Carmen is a hard-working and enterprising woman, who is 34 years old and the proprietor of a barber shop. She started it three years ago with a small chair and precarious conditions. Now she has her own premises, with two employees and better conditions. She will invest the loan in materials, another barber chair, scissors, Gillette, gel, shaving foam, etc.
	Max	40	Princess Unique works hard to support her family. Princess Unique is single and has a food vending business in the Philippines. Princess Unique requested a 5 PHP loan through NWTF to buy ingredients for her food production. Princess Unique has been in this business for one year now and also earns additional income from clothing sales. In the future, Princess Unique would like to save money to expand her business. She is a **transgender** client.

Annexe C : Liste des pays des emprunteurs et nombre de demandes de prêts par pays

Pays	N	Pays	N	Pays	N	Pays	N	Pays	N
Philippines	137474 (15,43%)	Honduras	10621 (1,19%)	Timor-Leste	4462 (0,5%)	Népal	1114 (0,13%)	Bélize	98 (0,01%)
Kenya	99772 (11,2%)	Inde	10438 (1,17%)	États-Unis	4443 (0,5%)	Zambie	1097 (0,12%)	Bosnie-Herzégovine	93 (0,01%)
Pérou	45043 (5,05%)	Ghana	9870 (1,11%)	Mozambique	4139 (0,46%)	Îles Salomon	994 (0,11%)	Porto Rico	77 (0,01%)
Salvador	43827 (4,92%)	Kirghizistan	9803 (1,1%)	Haïti	3857 (0,43%)	Laos	916 (0,1%)	Cote D'Ivoire	56 (0,01%)
Cambodge	42720 (4,79%)	Guatemala	9389 (1,05%)	Birmanie	3628 (0,41%)	Fiji	888 (0,1%)	Namibie	26 (0%)
Tadjikistan	31808 (3,57%)	Mali	9237 (1,04%)	Burkina Faso	3447 (0,39%)	Brésil	886 (0,1%)	Saint-Vincent-et-les-Grenadines	16 (0%)
Pakistan	31424 (3,53%)	Mexique	8713 (0,98%)	Costa Rica	3211 (0,36%)	Afghanistan	848 (0,1%)	Bulgarie	13 (0%)
Ouganda	29645 (3,33%)	Sénégal	8195 (0,92%)	Cameroun	3002 (0,34%)	Chili	764 (0,09%)	Papouasie-Nouvelle-Guinée	11 (0%)
Colombie	28795 (3,23%)	Jordanie	8029 (0,9%)	Albanie	2880 (0,32%)	Moldavie	761 (0,09%)	Vanuatu	3 (0%)
Équateur	27806 (3,12%)	Nigeria	7907 (0,89%)	Ukraine	2829 (0,32%)	Congo	640 (0,07%)	Guam	2 (0%)
Nicaragua	20787 (2,33%)	Sierra Leone	7874 (0,88%)	Bénin	2640 (0,3%)	Tonga	528 (0,06%)	Canada	2 (0%)
Paraguay	19524 (2,19%)	Togo	7359 (0,83%)	Égypte	2624 (0,29%)	Israël	426 (0,05%)	Bhoutan	2 (0%)
Liban	18025 (2,02%)	Zimbabwe	6142 (0,69%)	Yémen	2509 (0,28%)	L'Afrique du Sud	421 (0,05%)	Uruguay	1 (0%)
Vietnam	17815 (2%)	Indonésie	5761 (0,65%)	Irak	2506 (0,28%)	Lesotho	393 (0,04%)	Botswana	1 (0%)

Pays	N	Pays	N	Pays	N	Pays	N	Pays	N
Bolivie	17470 (1,96%)	Liberia	5627 (0,63%)	République dominicaine	2322 (0,26%)	Thaïlande	341 (0,04%)	Mauritanie	1 (0%)
Rwanda	12260 (1,38%)	Géorgie	5501 (0,62%)	Kosovo	1962 (0,22%)	Suriname	274 (0,03%)	Bangladesh	1 (0%)
Arménie	12226 (1,37%)	Mongolie	5164 (0,58%)	Malawi	1957 (0,22%)	Sri Lanka	224 (0,03%)		
Samoa	11901 (1,34%)	RDC	5057 (0,57%)	Turquie	1743 (0,2%)	Panama	151 (0,02%)		
Tanzanie	11602 (1,3%)	Madagascar	5015 (0,56%)	Soudan du Sud	1682 (0,19%)	Somalie	119 (0,01%)		
Palestine	11059 (1,24%)	Azerbaijan	4742 (0,53%)	Burundi	1598 (0,18%)	Chine	115 (0,01%)		

Annexe D : Coefficients de corrélation

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Duree	1																
Profit	-0,012	1															
Risque	0,004	-0,027	1														
Diversite	0,009	-0,080	0,018	1													
Interet_hunain	0,092	0,010	-0,021	-0,045	1												
Intrinseque	0,071	-0,053	-0,001	0,715	0,666	1											
Extrinseque	-0,006	0,705	0,690	-0,045	-0,008	-0,039	1										
Montant_pret	0,127	-0,121	0,035	0,048	-0,040	0,008	-0,063	1									
Risque_IMF	-0,007	0,088	0	-0,036	-0,051	-0,063	0,064	-0,042	1								
Duree_pret	0,279	0,032	0,012	-0,010	0,151	0,098	0,031	0,060	0,011	1							
Tenacite	0,024	-0,116	0,011	-0,057	0,026	-0,025	-0,076	0,003	0,060	-0,051	1						
Nivellement	0,026	-0,096	0,034	0,017	0,017	0,025	-0,045	0,126	-0,084	-0,001	-0,024	1					
Blame	-0,054	0,038	-0,011	-0,045	-0,055	-0,072	0,019	-0,076	0,229	-0,070	-0,072	-0,113	1				
Accomplishment	-0,024	0,141	-0,005	0,007	0,076	0,059	0,098	-0,048	0,083	-0,062	-0,032	-0,138	0,163	1			
Preo_actuelle	0,050	0,092	0,032	0,022	0,239	0,184	0,090	-0,019	-0,081	0,128	-0,156	0,056	-0,108	-0,099	1		
Caractere_concret	-0,010	-0,016	-0,044	0,026	-0,103	-0,053	-0,043	0,027	-0,044	-0,063	-0,101	-0,054	-0,010	-0,031	-0,069	1	
Variete	-0,053	0,070	-0,006	0,004	-0,053	-0,034	0,047	-0,136	0,080	-0,060	-0,052	-0,189	0,114	0,126	0,024	0,121	1

Toutes les corrélations supérieures à 0,01 en valeur absolue sont significatives au seuil de 1%

Annexe E : Résultats de MCO pour l'échantillon Éch3b

	2006-2019		2006-2009		2010-2019	
Variables	Modèle 1	Modèle 2	Modèle 3	Modèle 4	Modèle 5	Modèle 6
Variables indépendantes						
Intrinseque		0(0)		-0,020***(0,001)		0,009***(0,001)
Extrinseque		0(0)		0,009**(0,003)		-0,002(0,001)
Variables de contrôle						
Log (Montant_pret)	0,383***(0,003)	0,383***(0,003)		0,405***(0,007)	0,390***(0,003)	0,390***(0,003)
Risque_IMF	0,025***(0,001)	0,025***(0,001)		-0,025***(0,004)	0,023***(0,001)	0,023***(0,001)
Duree_pret	0,019***(0)	0,019***(0)		0,024***(0,001)	0,018***(0)	0,018***(0)
Accomplissement	0***(0)	0***(0)	0,002***(0)	0,002***(0)	0+(0)	0(0)
Blame	-0,010***(0,001)	-0,010***(0,001)	-0,042***(0,002)	-0,043***(0,002)	-0,006***(0,001)	-0,006***(0,001)
Nivellement	-0,005***(0)	-0,005***(0)	-0,005***(0,001)	-0,005***(0,001)	-0,005***(0)	-0,005***(0)
Tenacite	0,002***(0)	0,002***(0)	-0,001***(0)	-0,001***(0)	0,004***(0)	0,004***(0)
Variete	0,341***(0,028)	0,341***(0,028)	0,256***(0,001)	0,278***(0,060)	0,217***(0,032)	0,217***(0,032)
Caractere_concret	0,001***(0)	0,001***(0)	0,002***(0)	0,002***(0)	0,001***(0)	0,001***(0)
Preo_actuelle	-0,003***(0)	-0,003***(0)	-0,005***(0,001)	-0,004***(0)	0,001***(0)	0,001***(0)
Couverture						
Secteur						
Pays						
Constante	-1,640***(0,051)	-1,640***(0,051)	-1,558***(0,079)	-1,589***(0,079)	-2,676***(0,113)	-2,682***(0,113)
R^2	0,1702	0,1702	0,1983	0,1986	0,1589	0,159
DDL	108	110	72	74	106	108
DDL (résidus)	507 027	507 025	115 819	115 819	391 135	391 133

Annexe F : Résultats de l'algorithme génétique

	Perceptron multicouche		Perceptron multicouche	MLCT profond		MLCT profond
	Groupe	Individu	Score de propension	Groupe	Individu	Score de propension
Nombre de mots	---	---	---	72	115	340
Nombre de neurones (Couche 1)	22	46	20	16	67	75
Nombre de neurones (Couche 2)	44	23	34	33	36	61
Nombre de neurones (Couche 3)	11	37	19	18	16	30
Nombre de neurones (Couche 4)	37	18	8	35	22	57
Nombre de neurones (Couche 5)	25	14	---	21	26	48
Nombre de neurones (Couche 6)	2	3	---	17	28	10
Nombre de neurones (Couche 7)	---	---	---	20	11	---
Nombre de neurones (Couche 8)	----	---	---	5	20	---