A Work Project, presented as part of the requirements for the Award of a Master Degree in Management from the NOVA – School of Business and Economics

Conceptualising a Consumer Data Intelligence Service in perspective of the European Union's PSD2

Maurits Friso Cleveringa

Student Number: 25285

A Project carried out on the Master in Management Program, under the supervision of:

Miguel Muñoz Duarte

May 26th, 2017

Conceptualising a Consumer Data Intelligence Service in perspective of the European Union's PSD2

Abstract

In recent years the retail payments market has experienced significant technical advancement, with compelling growth in regard to development to new forms of payment and numerous payment services in the market place. In 2015, the European Commission (EC) introduced a new directive for payment services called PSD2, entering into force on 13 January 2018, to reduce barriers of entry and integrating retail payments across the European Union (EU). By granting third party non-financial institutions access to consumers' financial data, the EC is opening the financial services industry and levelling the playing field. This allows for a strong drive for innovation in the sector and for start-ups to offer new services to compete with the traditional banks. Consumer knowledge gained by the analysis of financial data becomes a strategic asset and allow businesses to connect with customers in new ways. New data intelligence services emerge in the financial technology industry seeking to close the knowledge gap between the expectations of customer experiences and currently available bank offerings. Therefore, the work project identifies how to provide valuable customer intelligence to businesses in the financial payments industry with regard to PSD2. To examine the new and emerging opportunities, we conceptualise a Consumer Data Intelligence Service, including statistical analyses to predict potential consumer behaviour. The proposed Consumer Data Intelligence Service concept suggests that turning relevant customer data into valuable insights through data analytics will allow businesses to act on the new opportunities.

"I would like to assert that data will be the basis of competitive advantage for any organisation that you run." – Ginni Rometty, CEO of IBM (7.03.2013, Council on foreign relations – Corporate Conference)

Keywords: Consumer Data Intelligence, PSD2, predictive analysis, customer insights

Introduction

In recent years access to data and its use for commercial purposes has gathered extensive attention as the digitalisation of the world progresses. The enterprise value of data has risen to new heights and become more valuable than ever before (Disparte et al., 2016). In today's digital era, data is becoming the new commodity seeding a rapidly growing industry which is making companies such as Alphabet, Facebook and Amazon seem to be omnipotent (Economist, 2017). Such data driven companies are pushing into numerous industries catering to consumers' and business needs. One of which is the new information technology driven payment industry in the EU.

Taking into consideration the fast advancements in technology driving consumers' expectations and imagination in payment industry, new product offerings emerge. One major challenge in designing new technological products and services in the EU is to comply with the current guidelines. The EC and the European Banking Authority (EBA) have been advocating the introduction of principles to monitor and oversee the industry through Payment Service Directives (PSD). Before the introduction of the EU's first PSD in 2007 the state of payment regulations across the EU varied from country to country. The goal of the PSD was to establish a modern and coherent legal framework for payment services among the then 27 member states defining the new payment services industry (European Commission, 2007). The provision allowed non-bank financial institutions to provide financial services to consumers and thus remove a barrier of entry to the market posed by regulation.

To take into account new services and players, to address competition in the market and to broaden the scope of the PSD the EC decided to review the Directive in 2013. The main stakeholders addressed are financial institutions, such as banks and credit unions, and the financial technology industry containing start-ups and technology firms. Besides providing payment services, new providers can also consolidate one's account information in one place and the new Directive (PSD2) calls for banks to open up access to their accounts' data to third parties through Application Programming Interfaces (APIs). This allows Third Party Providers (TPP) involved in financial technology to acquire data about financial institutions' customers. Hereby, valuable insights for businesses and end customers can be collected. PSD2 will enter into force on 13 January 2018. The goal is to stimulate competition by removing barriers resulting in innovative and cheaper payment solutions for consumers in Europe. Incumbent banks will be forced to adapt to a new market environment as new opportunities arise. In this context, data intelligence services will find lucrative business prospects to cater to Account Information Service Providers (AISP) and Payment Initiation Services Providers (PISP). AISPs allow to centralise customer account information and analyse consumer behaviour such as spending patterns, expenses and financial needs (Evry, 2016). PISPs, instead, facilitate user payments and provide a bridge between the consumer and retailers or financial offering.

The conceptualisation and structure of a Consumer Data Intelligence Service (CDIS) is the topic of this paper. At this stage the theorised service is a proposal and does not include a comprehensive business model in form of a business model canvas. However, this work project serves as a basis for business models that can be ventured in the future. It is up to the creativity and risk appetite of entrepreneurs to start a venture which follows the proposed conceptual architecture and examined analyses.

The main research question of this paper is to assess how a CDIS provides new business insights to TPPs, AISPs and PISPs using the newly accessible banking data of European customers. To answer this question, we introduce a conceptual architecture and use a data set containing a fictional sample of 1000 credit customers from Germany to predict credit scoring.

The remaining structure of this paper is organised as follows. First, the service context and business background is looked at. Second, related work to advanced statistical methods, consumer intelligence, and PSD2 is reviewed. Third, the conceptual architecture of the proposed CDIS is described. Fourth, we continue with the description of the methodology of our statistical analysis and its results. Lastly, the paper will critically reflect on the limitations of this work project and conclude with future implications for the data analytics ecosystem and in research.

Service Concept

Consumer Data Intelligence Service

Through the aforementioned new payment Directives, the positions of financial institutions will be challenged by new entrants. Incumbent players face an uncertain future in light of what is to change the access to their customers' data. Banks in the United Kingdom are expected to lose revenues of retail payments of up to 43% according to the whitepaper of Accenture (2016). One strategic option includes becoming a utility type service for digital payments and account information. In this case, they risk losing engagement with the customer to build loyalty and deliver a rewarding customer experience. A second strategic option for the involved stakeholders is to become a commonplace bank present in customers' everyday lives. Here, a bank is bound to expand its ecosystem and aggregate value to grow new revenue streams and apprehend customer ownership. There is thus a great challenge for the financial institutions coming up. Furthermore, innovative business ideas will enter the market such as the proposed CDIS. In this section the service background and conceptual design will be introduced.

The author's interest in financial payments stems from initial professional experience during an internship in a large telecommunications company. While mobile services were originally in focus, it was the launch of a new mobile payment product that captivated most attention. Here, details of complex intermediary brokering processes with merchants and acquirers who process payments, data ownership difficulties, and hurdles involved in the establishment of a payment service became clear. The barriers to entry for a start-up, and even for a large telecommunications company, were evidently hard to surpass. Noticeable barriers are knowledge of customers and access to their financial data. Limited customer data in form of personal information could be gathered by the consumers' use of the service and through granting permissions by accepting the terms & conditions of the service. Nevertheless, because the company is in the telecommunications business, it has access to large amounts of data about the customers who use other services such as mobile phone plans. This dominant position of valuable access to data and financial information allows the company to operate on its own by offering new financial services. The same is present in financial institutions such as banks and credit unions. Here, financial institutions hold an asset in their hands in a monopolistic manner. This is due to various reasons according to Malgieri (2016). Firstly, the data is becoming of great value as it presents unique opportunities to market tailored products and services to their customers. Secondly, some of the data is subject to intellectual property rights and regarded as trade secrets. Lastly, because of security concerns and data protection. The data may be manipulated or stolen and exposed for fraudulent and criminal purposes.

Besides also imposing new security requirements augmenting the security levels for payment processing and TPPs, PSD2 disrupts the monopolistic ownership of customer data and pushes to share the information with more businesses. Hence, a new concept for a service is presented.

The suggested CDIS is based on the following assumptions aimed at clarifying that this service is a conceptualisation. First, even though the PSD2 contains more components than only the accessibility of financial data, its wider scope will be the main emphasis of this work project. The exact rules and regulations on which data will be shared is yet to be specified by the EBA. Second, this work project assumes that the obtainable data from bank customers will reflect the data included in the data set analysed in a later section. Third, the data analysis serves as a base to develop classification methods that can contribute to increase the accuracy of predicting consumer behaviour to deliver customer insights a service.

Related work

This section draws a review from statistical analysis theory, consumer intelligence literature and related work on PSD2 and its impact. It contributes to the understanding of the theoretical background of this work paper.

Business Intelligence

Business Intelligence is the process of delivering actionable business decisions from analytical manipulation and presentation of data within the business environment (Sherif, 2016). This strand of related literature comprises Customer Intelligence research. No particular tool covers Business or Consumer Intelligence as the tool serves as the delivery mechanism for the business logic. The particular value of the knowledge acquired through Consumer Intelligence stems from the value acting on those insights.

Related Work on PSD2

The implications of PSD2 in regard to the EC's Digital Market Strategy (in force since May 2015) have been analysed by Donnelly (2016) and resulted in the identification of gaps in the framework. According to Donnelly (2016), the EBA will have a major effect on whether the directive and the Digital Market Strategy will be harmonised. Furthermore, a study was performed to explore how the introduction of new regulations influence institutional complexity of businesses (Haataja, 2015). Hereby, it was found that the changes in market conditions allow for vast new possibilities of partnerships and integration for prospective firms. Existing companies may see the new directive as a threat as new entrants stimulate further competition.

Advanced statistical methods for data analysis

The field of research on statistical analysis comprises multiple disciplines. As technology advancements over the past decades allowed to process larger amounts of data in shorter periods of time, more advanced techniques were developed. These include probabilistic classification techniques in the field of predictive analytics. Hereby, predictive modelling is the method which defines the process of developing a way of understanding and quantifying prediction models of future data with accuracy (Kuhn, 2013). The work of Kuhn and Johnson serves to thoroughly understand the principles of Random Forests, Naïve Bayes classifiers, and Support Vector Machines. As discussed by Breiman (1999) classification accuracy can be significantly improved by growing an ensemble of decision trees and letting them determine the most important class. Naïve Bayes classifiers examine the probability of the outcome of a class being predicted correctly based on previously observed predictors (Kuhn, 2013). It does, though, make a strong assumption of all predictors being independent. This institutes a trade-off between complexity of calculations and simplifying the predictor probabilities. Support Vector Machines are a learning machine for two group classification problems where the margin, the distance between the classification boundary and the closest training set point, judges the efficacy of a model (Vapnik, 1995).

Research Question

The author’s work focuses on the conceptualisation of a service to cater customer intelligence data to TPPs in form of AISPs and PISPs. The following research question is investigated: With regard to the European Union’s Payment Service Directive 2, which enters into force on 13 January 2018, how can a Consumer Data Intelligence Service provide valuable customer intelligence to new businesses in the financial payments industry?

The following section provides the conceptual architecture of a CDIS. A data workflow process matched with the examination of statistical analyses complements the concept. Additionally, three predictive models are evaluated to demonstrate potential data intelligence.

Consumer Intelligence

Through the power of knowledge of any specific consumer, consumer segments will cede to cater to the individual as a segment (IBM, 2016). Technology shifts will change the way businesses deliver value. Rather than meeting the needs of different consumer segments such as demographic or psychographic segmentation, businesses will be able to truly serve the individual with highly tailored products and services. "*What you will see with rapid data and social sharing is the death of the average and the era of you*," Ginni Rometty, CEO of IBM (2013).

Conceptual Architecture

The following figure illustrates the key elements of CDIS and the directly involved stakeholders:

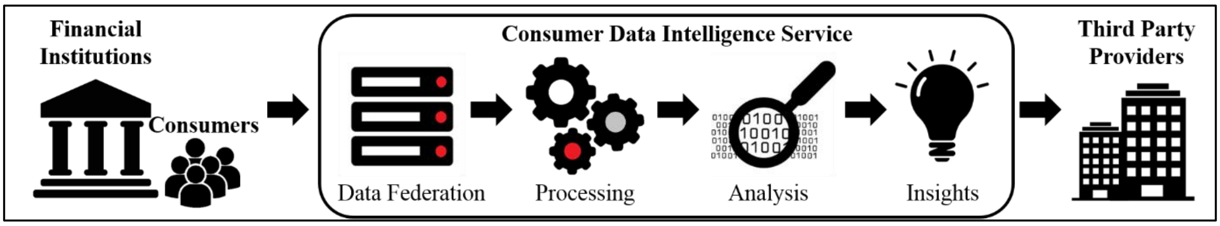


Figure 1, Conceptual Architecture

The CDIS is the connecting element between financial institutions' customer data and TPP such as AISPs and PISPs. As visualised in Figure 1 the four main elements are Data Federation, Data Processing, Data Analysis and Insights. Hence, to successfully deliver a CDIS a close collaboration with both financial institutions and TPPs is necessary.

According to Accenture (2016), in the market environment of the CDIS concept herewith proposed, four strategies emerge for future TPPs to gain a competitive advantage. First, TPPs will need to innovate and launch cutting-edge products to meet consumer expectations. Second, companies involved in the new market environment of the financial payment industry need to form strategic alliances to offer wide-ranging retail and financial products. Third, TPPs need to formulate clear data sharing strategies to gain trust and loyalty by consumers. Fourth, superior customer experience needs to be granted through added security.

One successful example of a payment service, which emerged in the Netherlands before the 2013 revision of PSD, is iDeal. A company founded in 2005 which today has 13 leading Dutch banks participating. Over the last five years, iDeal has become the largest e-commerce payment method in the Netherlands whilst also providing limited consumer intelligence insights to the participating banks. Over the past 5 years iDeal experienced an average annual growth of 26.52% (iDeal, 2017).

Methodology

This work project serves as a concept for new TPP businesses to be founded or modify the scope of existing ones. The business concept contains theory involving prediction models to engage in sharing banking insights to the CDIS' customers. Hereby, the prediction models focus on predicting credit scores of bank customers through data sources such as the newly available banking data. In order to comprehend the complete system of data management to predict credit scores we need to understand the steps involved from start to finish.

We begin by looking at the following workflow process in Figure 2.

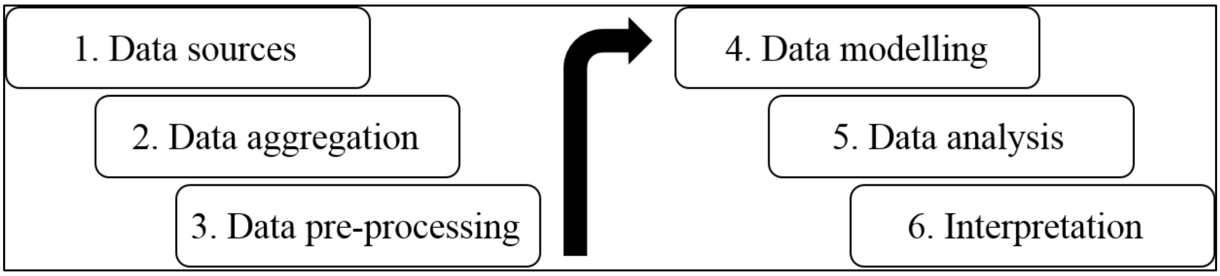


Figure 2, Data Workflow Process.

The data workflow process consist of six steps which will be described in the subsequent section. According to Pyle (1999) we explore data with the goal of discovering information and understanding something about the real world through objects. Identifiable objects are rich, dynamic, and complex in features which can be defined and measured. Here, many data sources exist as measurements which can be taken in various ways. These sources need to be identified and the data needs to be gathered and collected in databases. Once the sources have been recognised the process of data aggregation begins. This process involves mining and searching databases to be able to compile information with the aim of preparing data sets for data processing (Pyle, 1999). Nevertheless, the quality of data sets can be inferior due to irrelevant and superfluous information. This can result in ambiguous analyses as data sets could represent a mistaken study or fail to discover anything useful at all. Hence, pre-processing is a necessary step to achieve the best analytical performance of the data set. The techniques involved are amongst others data cleaning, normalization, extraction, and reduction. They solve several types of problems including noisy data, redundant data, missing data values, or the excessive size of the data set. Considerable amounts of time are spent on this process (Kotsiantis et al., 2006). The product after the data has been pre-processed is the basis for us to build a model. According to Kuhn (2013), there are two main steps in the model building process. Firstly, as we want to predict the credit score of a consumer we need to understand the characteristics and relationships of our predictors. Secondly, we create a model using older data and test it on the newer data. Thus, we have a training set and a test set. The test set is used for validation purposes (Kuhn, 2013).

Data analysis

A very important part of the CDIS and its Data Workflow Process is the Data analysis. First, we examine the data set and its measures and dimensions. Second, we summarise the descriptive statistics of our observations. Third, we evaluate the selection of statistical models for our analysis and inspect the theory behind them.

Data description

The data set at hand contains information of 1000 observations of credit takers from a fictional financial institution in Germany. The sample is publicly available and contains a subset of variables containing information about the credit class, the credit taken (Duration in months, Amount in Euros, Instalment rate percentage), the credit history (Number of existing credits, Number of previous credits, Credit payment punctuality, Credit purpose, Other instalment plans, Other debtors), and personal information (Age, Employment duration, Type of employment, Type of assets owned, Type of housing). Hereby, we use several binary variables instead of categorical variables to ensure mutual exclusivity. A value of 0 in a binary variable will cause this value to not influence the classification in any way. The credit class explains the credit score which ranges from 0 to 10, where 10 is the best score and any value lower than 6 depicts a bad score. All variable categories can be found in Appendix 3.

Descriptive statistics

To form a basis of our quantitative analysis we look at the elementary features of the data in our study. Hereby, we summarise the data to allow us to make simple interpretation of the sample. They form the basis of the quantitative analysis of the data set. However, even though patterns might emerge, no conclusions regarding any hypothesis can be reached as the statistics are not conclusive.

In Appendix 2 we summarise the main descriptive statistics of the sample. Our sample contains consumers who are on average 35.5 years old and range in the ages 19 to 75 years old. The unemployment rate in our sample is just over 6% which is just slightly higher that the unemployment rate of 5.8% in Germany in April 2017 (BA, 2017). Almost 78% of credit takers are skilled employees or possess high qualifications. The average credit score is 6.6 and among all credit takers 70% have a good credit score. Hereby, the average amount of the credit is approximately € 3270 and the average duration is just under 21 months. Only 4% of our sample hasn’t taken any credit in the past and approximately 60% have duly paid all past credits. The purpose most credit takers indicated to use the credit for is to buy a new car.

Selection of statistical models for analyses

Choosing a statistical method for a predictive analysis is a comprehensive task. There is no singular best model to achieve the best results as the nature of the problem, the prediction of credit scores, and the structure of the data we are working with need to be considered. As described before, our data set contains mostly categorical variables and the amount of data is small in size. We want to define the correct credit score for each customer and will use a classification method. Based on the present information we will make use of the following three main groups of classification methods: Random forests, Naïve Bayes algorithms, and Support Vector Machines.

In our analysis we want to predict with accuracy the future behaviour of a user in regards to their credit score. We create three models using two sets of data, a training set and a test set (Finlay, 2014). The training set is composed of 800 entries of the data set while we test the model on the remaining 200.

Random Forests

Random forests are a combination of tree predictors in which each tree depends on the values of a random vector tested independently and with the same distribution for all trees in the forest (Breiman et al., 1999). Prediction trees are a dependence technique where the objective is to see how some independent variables help to predict an outcome variable (Ho, 1995). They are a classification method consisting of a collection classifiers structured in trees.

Naïve Bayes classification

The probabilistic model of Naïve Bayes classifiers is based on Bayes' theorem which assumes that all dimensions in a dataset are mutually independent. In practice, the independence assumption cannot always be realised, but naive Bayes classifiers still tend to perform very well under this unrealistic assumption (Russell, 2010). Especially for small sample sizes or high dimensionality, Naïve Bayes classifiers can perform well and with significance. The model follows the following base equation:

$$\mathrm{P}(\,\mathrm{c} \mid \mathrm{x}\,) = \frac{\mathrm{P}(\mathrm{x} \mid \mathrm{c}) \cdot \mathrm{P}(\mathrm{c})}{\mathrm{P}(\mathrm{x})}$$

Equation 1, Naïve Bayes theorem

Support Vector Machines

A Support Vector Machine (SVM) is a classification method which is defined by discriminating classes by a separating hyperplane. An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. In other words, given labelled training data (supervised learning), the algorithm outputs an optimal hyperplane which categorises new data observations (Bellotti, 2009).

Results

Having performed the analysis on the three classification methods we can review the models' results. First, the levels of accuracy of the model are being examined. Hereby, we choose the model that best predicts the consumers' credit score. Second, we investigate the Receiver Operating Characteristic Curve (ROC) including the Area under Curve (AUC). Third, the variable importance is evaluated as we identify the most significant variables to increase the predictive accuracy of the models.

Best model accuracy

The best ranking model is constituted by the highest level of accuracy for classifying credit scores in the present data set. From the three analyses performed the Random Forest analysis scored the highest level of accuracy with a degree of 77.5%. The classification accuracy of the Support Vector Machine lies in the same range at 77%. Both scores are better than the Naïve Bayes classifier with an accuracy of 72%.

Receiver Operating Characteristic and Area under Curve

To compare the models, the one with the highest area under the curve can be seen as the best model. We plotted an ROC curve for all three models (see Appendix 4). Plotting the true positive rate (TPR) against the false positive rate (FPR) at various threshold settings across the model creates the ROC curve. Hereby, the TPR represents the sensitivity or probability of detection of the predictor. The FPR, instead, shows the probability of a false alarm. Thus, the sensitivity as a function of false alarm is the ROC. Since the ROC of the models can be visibly similar it is best to look at the area of the respective functions.

The Area under the Curve (AUC) serves the purpose of better comparing different models. The model with the highest area under the curve is seen as the best.

	Random Forest	*Naïve Bayes*	*Support Vector Machine*
Area under Curve	0.7733813	0.7522113	0.7654205

Figure 3: AUC across three models

Variable importance

Variable Importance ranks the variables according to their ability to reduce the predictive accuracy of the model if removed from the model. The higher the score the more important the variables become to predict the credit score. Random Forest analysis has a built-in variable importance score. Variable importance is automatically scaled to be between 0 and 100 whereas 100 is the most important factor.

In Appendix 5 we depict a summary of ranking of the most important variables across the three classification methods. As with the accuracy of prediction power and the ROC curves the results did not vary to a great extent. In all three cases the variables Checking Account Status, Duration, Amount, and Checking Account Negative were ranked among the four most important variables with Checking Account Status being the most important every time.

Limitations and Research

The Consumer Data Intelligence Service proposed in this work project experiences a few limitations. First of all, regulatory challenges need to be overcome as the exact details of the implementation of PSD2 still need to be set in place by the EU member states. Second, we rely on a fictitious data set which is not representative for the population. Lastly, two major stakeholders could not be involved in the work project as a collaboration with banks and TPPs could not be realised within the scope of this work project.

Rules and Guidelines

The European Banking Authority (EBA), together with all involved stakeholders, including the payment services market, needs to develop and confirm several technical regulation standards and guidelines balanced between pros and cons. Regulation which is too specific might hinder future innovation whereas a lack of clarity could make incumbent financial institutions more reluctant on opening access. Either way, it is crucial for regulators -preferably in consultation with the diverse stakeholders in order to enhance acceptance by the finance industry- to find the right balance between these two in the context of PSD2. Nevertheless, it is likely that some issues may need to be readdressed after the initial implementation stages as needs for modifications will only surface once the Directive takes effect. Experiences from the review process of the first PSD refer.

Real Data availability

The data set used to evaluate credit scores is fictitious and not derived from real financial transactions. Hence, it is imperative to understand that the analyses performed are a demonstration of what can be achieved by a consumer insights service company. In the near future, banks need to comply with the national regulations established by the EU member governments and provide APIs access the data.

Stakeholder collaboration

Due to the early stage of conceptualisation of CDIS, and to the banks' understandable reticence to share real data, unfortunately, no bank collaboration could be established to apply and refine the concept with a real company or bank using their data. In preparation for the removal of barriers by the new Directive, Nordea bank from Sweden is launching an "open banking portal" in summer 2017 to give TPPs access to their APIs (Haaramo, 2017; Nordea, 2017). Hereby, a dialogue with TPPs will be formed and a sandbox environment for developers to test their products and services will be introduced.

Further research

Further research would be needed to investigate the strategic options financial institutions choose to pursue and measure their success, including key decision-making metrics. Furthermore, the impact of newly obtained customer insights on the product offering as well as on the level of competition within the financial payment industry should be studied. Lastly, and in view to support the revision of the PSD2 in order to enhance its effectiveness, pre-competition cooperation between stakeholders merits future research.

Conclusion

The digital revolution is progressing the human-machine interaction every day as more and more is driven by computerisation and digitalisation. The digitalised world is reaching into every corner of our lives. Business dynamics are changing and industries need to adapt to new challenges posed by advancements in technology and deliver value to fulfil consumer expectations. The financial payment industry in the EU is facing a major opportunity with the introduction of PSD2. The deregulation of access to financial data will have a major impact on the world of digital payments and retail banking. New opportunities arise and competition in the financial technology sector will be stimulated and broadened beyond the incumbent financial institutions to start-ups and technology firms. In this context, the work project provides a conceptualisation of a service, i.e. CDIS, aimed at providing consumer insights and answers the following question: How can a Consumer Data Intelligence Service provide valuable customer intelligence to new businesses in the financial payments industry? We reviewed associated work by looking at literature from the domains of statistical analysis, business intelligence, and PSD2. An overview of consumer insight opportunities with regard to PSD2 was introduced proposing the conceptual architecture of a CDIS. Hereby, we identify the role of the CDIS as a provider of consumer insights for TPPs and existing technology firms. The service is based on data analytics which involve a data workflow processes and statistical analysis comprising three predictive models. The results suggest the Random Forest model to be the best in predicting credit scores.

In conclusion, the work project convincingly demonstrates the organisation of a CDIS and its ability to accurately predict consumer behaviour for consumer intelligence purposes.

References

Bundesagentur für Arbeit. 2017. “Arbeitslosigkeit, Unterbeschäftigung und gemeldetes Stellenangebot”. Accessed May 15, 2017. https://statistik.arbeitsagentur.de/Navigation/Statistik/Statistik-nach-Themen/Arbeitslose-und-gemeldetes-Stellenangebot/Arbeislose-und-gemeldetes-Stellenangebot-Nav.html

Bellotti, T. & Crook, J. 2009. “Support vector machines for credit scoring and discovery of significant features”. *Expert Systems with Applications*, 36 (2).

Breiman, L. 1996. "Bagging predictors". *Machine Learning*, 24 (2), 123–140.

The Economist. 2017. “The world’s most valuable resource is no longer oil, but data”. Accessed May 10, 2017. http://www.economist.com/news/leaders/21721656-data-economy-demands-new-approach-antitrust-rules-worlds-most-valuable-resource

European Commission. 2007. “Payment Service Directive”. Directive 2007/64/EC of the European Parliament and of the Council.

European Commission. 2015. “Revision of the Payment Service Directive”. Directive EU 2015/2366 of the European Parliament and of the Council.

Disparte, D. & Wagner, D. 2016. “Do You Know What Your Company’s Data Is Worth?” *Harvard Business Review*. Accessed May 10, 2017. https://hbr.org/2016/09/do-you-know-what-your-companys-data-is-worth

Finlay, S. 2014. *Predictive Analytics, Data Mining and Big Data. Myths, Misconceptions and Methods*. Basingstoke: Palgrave Macmillan.

Rometty, G. 2013. “A Conversation with Ginni Rometty”. Accessed May 6, 2017. https://www.youtube.com/watch?v=SUoCHC-i7_o

Haaramo, E. 2017. “Nordea goes beyond what is mandatory for PSD2”. Accessed May 10, 2017. http://www.computerweekly.com/news/450414988/Nordea-goes-beyond-what-is-mandatory-for-PSD2

Hafstad, T., Hjort, G., Johansson, F., Crompton, D., Ullgren, J., Johnston, M., Oyna, M. 2016. “PSD2 – Strategic opportunities beyond compliance”. *Evry*.

Ho, T.K. 1995. “Random Decision Forests.” *IEEE Computer Society*, 278–282.

Kuhn, M. & Johnson, K. 2013. *Applied predictive modelling*. New York: Springer New York.

Light, J., McFarlane, A., Killian, B., Ilka, R. 2016. “Seizing the Opportunities Unlocked by the EU’s Revised Payment Services Directive”. *Accenture.*

Malgieri, G. 2016. "Ownership of customer data in the European Union: quasi-property as comparative solution?”. *Journal of Internet Law*, vol. 20, no. 5, pp. 3.

Nordea. 2017. ” Nordea prepares for PSD2 with first iteration of Open Banking portal.” Accessed May 17, 2017. https://www.nordea.com/en/our-services/cashmanagement/cash-management-news/cash-management-news-archive/2017/open-banking-portal.html

Russell, S.J. & Norvig, P. 2010. *Artificial intelligence: a modern approach*. Upper Saddle River: Prentice Hall

Salmony, M. 2014. “Access to accounts: Why banks should embrace an open future.” *Journal of payments strategy & systems*, 8(2), 157–171.

Sharma, D., Overstreet, G., Beling, P. 2009. “Enhancing Predictive capability of Credit Scoring Using Affordability Data”. *Casualty Actuarial Society Working Paper*.

Sherif, A. 2016. *Practical business intelligence: learn to get the most out of your business data to optimize your business*. Mumbai, Birmingham: Packt Publishing.

Tan, W., Fan, Y., Ghoneim, A., Hossain, A., Dustdar, S. 2016. "From the Service-Oriented Architecture to the Web API Economy." *IEEE Internet Computing,* 20(4), 64-68.

Vapnik, V. & Cortes, C. (1995)."Support-Vector Networks". *Machine Learning*, vol. 20, no. 3, pp. 273-297.

Wilcox, Rand R. 2010. *Fundamentals of Modern Statistical Methods.* New York: Springer, 200–213.

Appendices

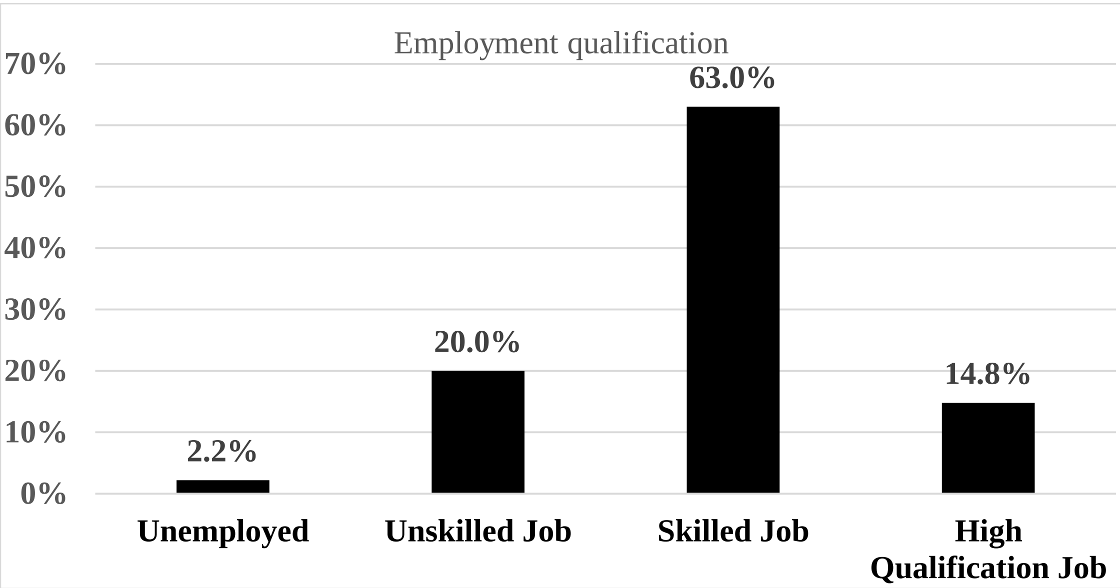
Appendix 1: List of acronyms and abbreviations

CDIS	Consumer Data Intelligence Service	EBA	European Banking Authority
PSD	Payment Service Directive	RF	Random Forests
TPP	Third Party Provider	NB	Naïve Bayes
AISP	Account Information Service Provider	SVM	Support Vector Machines
PISP	Payment Initiation Service Provider	ROC	Receiving Operator Curve
API	Application Programming Interface	AUC	Area under Curve
EU	European Union	TPR	True Positive Rate
EC	European Commission	FPR	False Positive Rate

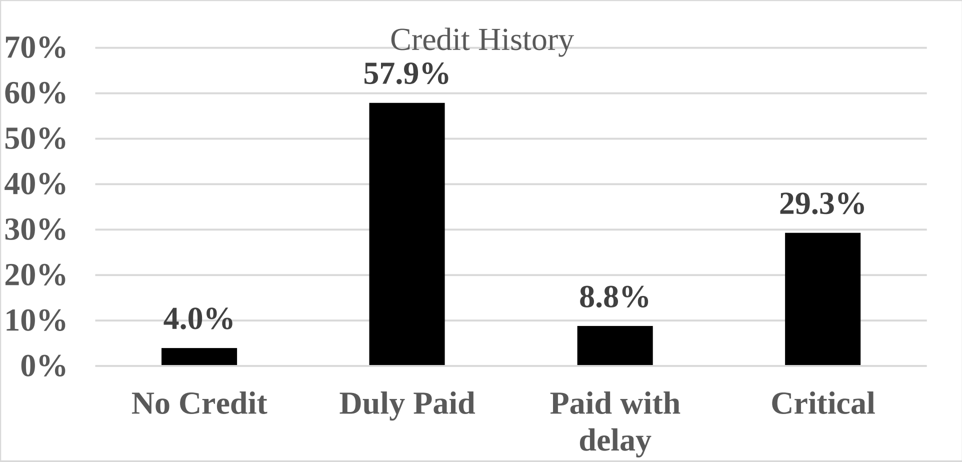
Appendix 2: Descriptive statistics of the sample

VARIABLE	MEAN	MIN	MAX
AGE	35.54	19	75
DURATION	20.9	4	72
AMOUNT	3271.26	250	18424
CREDIT SCORE	6.58	0	10

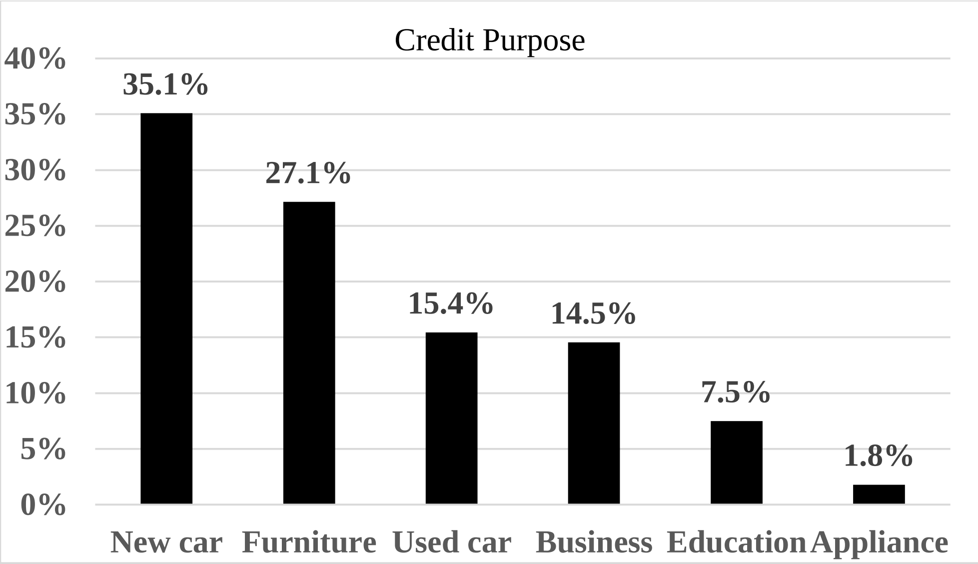
Further dimensions:



Employment qualifications



Credit History

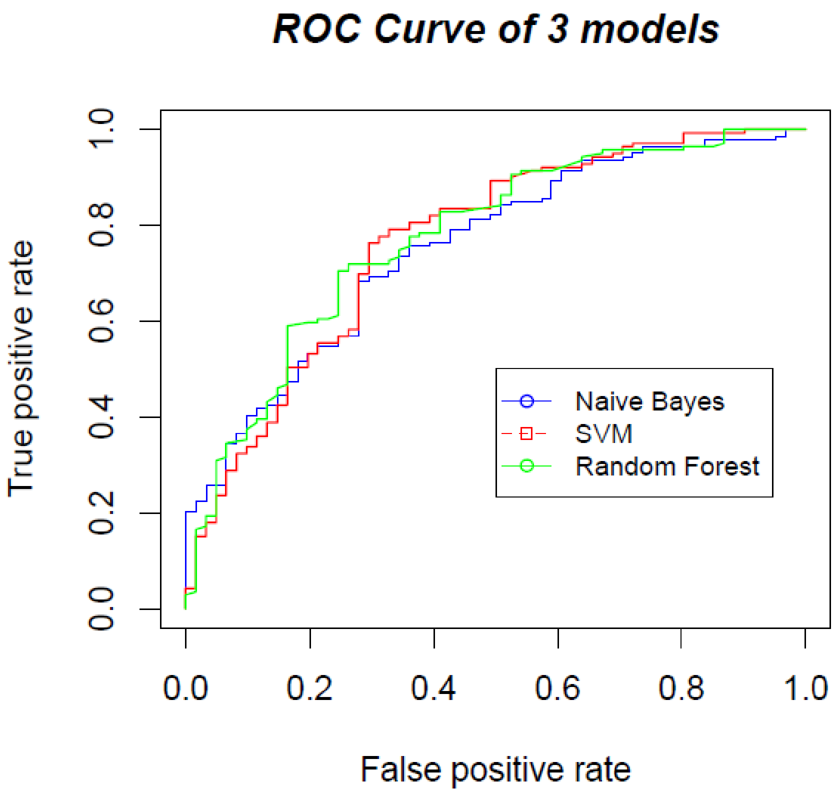


Credit Purpose

Appendix 3: Variable categories

Duration Amount Instalment Rate Residence Duration Age Number of Credits	Telephone Credit class Checking Account Status Credit History Credit Purpose Employment Duration	Other Debtors Property Other Instalment Plans Housing Employment Qualification

Appendix 4: ROC of 3 models



Appendix 5: Ranking of the most important variables

Random Forests	Support Vector Machines	Naïve Bayes classifier
Checking Account Status	Checking Account Status	Checking Account Status
Duration	Duration	Duration
Amount	Checking Account Negative	Checking Account Negative
Checking Account Negative	Critical Credit History	Critical Credit History
Age	Checking Account <200.000€	Checking Account <200.000€