UNIVERSIDADE DA BEIRA INTERIOR
Engenharia

Unsupervised and Language Independent Approach to Extremism and Collective Radicalization Understanding

Miguel Ângelo Serra Albardeiro

Dissertação para obtenção do Grau de Mestre em

Engenharia Informática

(2º ciclo de estudos)

Orientador: Prof. Doutor Sebastião Pais
Co-orientador: Prof. Doutor João Paulo Cordeiro

Covilhã, Dezembro de 2019

Dedicatória

Dedico este trabalho a todos o que me deram o apoio que eu precisava para cumprir esta etapa do meu percurso académico, em especial aos meus pais e à Sabrina.

Agradecimentos

Este trabalho foi suportado por financiamento nacional pela FCT Fundacão para a Ciência e a Tecnologia, através do projeto MOVES - PTDC/EEI-AUT/28918/2017.
Em primeiro lugar, quero muito agradecer ao meu orientador Professor Doutor Sebastião Pais e ao co-orientador Professor Doutor João Paulo Cordeiro, pela disponibilidade, conselhos e orientação em todo o desenrolar desta investigação. De seguida, quero agradecer a todo o corpo docente e não docente do departamento de informática por toda a ajuda sempre disponibilizada e ainda a todos os docentes de outros departamentos que transmitiram o seu conhecimento da melhor forma que sabiam, levando a que eu seguisse o percurso que me fez chegar a esta etapa. Não esquecendo o agradecimento especial a todos os membros do HULTIG.
Agradeço a todos os meus amigos que me apoiaram e que estiveram comigo não só nos momentos de diversão, mas que estiveram comigo também nas piores alturas. E, não posso deixar de agradecer à Sabrina que desde o início me incentivou a percorrer o caminho para um futuro brilhante. Que antes mesmo de eu acreditar, já ela acreditava que eu chegaria longe. Obrigado pela paciência nos momentos em que eu a perdia, pelo apoio incondicional, por me encorajar sempre a melhorar e por me proporcionar alguns dos melhores momentos da minha vida, mesmo neste percurso. Deixando também uma palavra de apreço aos seus pais que também eles deram o seu contributo para este percurso.
E, por último, mas mais importante, quero agradecer aos meus pais que fizeram enormes sacrifícios para que eu investisse na minha formação e que me deram a chance de ter um futuro melhor. A eles devo tudo o que sou e tudo o que tenho. Sempre foram pais presentes e muito dedicados, que sempre me colocaram em primeiro lugar. A eles agradeço a educação, a formação e acima de tudo a vida!
Um enorme obrigado a todos!
Miguel Albardeiro

Acknowledgements

This work was supported by National Founding from the FCT Fundacão para a Ciência e a Tecnologia, through the MOVES Project - PTDC/EEI-AUT/28918/2017.
First of all, I would like to thank my advisor Professor Sebastião Pais and co-advisor Professor João Paulo Cordeiro for their availability, advice and guidance throughout this research. Next, I want to thank all of the IT department's professors and staff for all the help they have always provided, and all the professors from other departments who conveyed their knowledge as best they knew, leading me to follow the route. that made me get to this stage. Not forgetting the special thanks to all members of HULTIG.
I thank all my friends who supported me and who were with me not only during the fun times but who was with me also at the worst times. I cannot help but thank Sabrina who, from the beginning, encouraged me to walk the path to a bright future. That even before I believed, she already believed that I would get far. Thank you for your patience at times when I lost it, for your unconditional support, for always encouraging me to improve and for providing me with some of the best moments of my life, even on this journey. Also leaving a word of appreciation to your parents that they too have contributed to this journey.
And last, but most importantly, I want to thank my parents who made huge sacrifices for me to invest in my education and gave me a chance to have a better future. I owe them everything I am and everything I have. They were always present and very dedicated parents, who always put me first place. I thank them for my education and above all my life!
Thank you all!
Miguel Albardeiro

Resumo

Cada vez mais nos social medias encontramos grupos que se organizam para protestarem contra algo e, muitas vezes, nesses mesmos grupos por vezes estão inseridos membros com ideologias extremistas, com o intuito de destabilizar a ordem publica e espalhar os seus ideias recorrendo ao terror. Verifica-se que estes casos são cada vez mais recorrentes, ao criar-se um grupo específico cuja finalidade é a realização de protestos pacíficos com objetivos liberais e concretos, existe muitas vezes alguém que inicia um tópico com linguagem extremista. E, daqui, justificado pela influência de grupo, é possível ter-se em consideração a possibilidade de radicalização coletiva.

O objetivo desta investigação é criar uma abordagem para deteção de casos de extremismo e radicalização coletiva em redes sociais e isto deve ser feito de forma não supervisionada e independente da língua.

Os métodos utilizados foram: a criação de um léxico de termos de sentimento extremo denominado ExtremeSentiLex e de um classificador de sentimentos extremos em que o input são os termos de sentimento extremo e os posts de redes sociais. Para o desenvolvimento destas ferramentas foram utilizados métodos de processamento da linguagem natural puramente estatísticos. Sendo que, para podermos validar o ExtremeSentiLex este foi aplicado recorrendo ao classificador de sentimentos extremos e aos posts de input que são analisados que são posts de datasets já validados pela comunidade cientifica.

Para um estudo comparativo, são utilizados word embeddings para expandir o ExtremeSentiLex obtido e é também feito um teste em que o ExtremeSentiLex é balanceado e aplicado a um dataset também balanceado a nível da polaridade de sentimentos.

Os resultados obtidos nesta investigação e que serão disponibilizados para a comunidade cientifica são: o ExtremeSentiLex e datasets, que foram avaliados, relativamente à presença de sentimentos extremos; Os testes efetuados aquando da validação do ExtremeSentiLex: o nível de precisão ao encontrar sentimentos extremos na polaridade correta foi muito elevada. Já aquando da aplicação dos word embeddings os resultados pioraram; Com ExtremeSentiLex e dataset balanceados, os resultados melhoraram.

Concluí-se que o ExtremeSentiLex é adequado para a deteção de sentimentos extremos em texto. Detetou-se ainda que com a ajuda de especialistas na área da linguística e da psicologia o ExtremeSentiLex poderia ser aprimorado. Contudo o objetivo desta investigação era apenas fazê-lo recorrendo a métodos puramente estatísticos.

Palavras-chave

Análise de Sentimentos, Análise de Sentimentos Extremos, Social Media, Redes Sociais, Processamento de Linguagem Natural, Processamento de Linguagem Natural Estatistico, Extremismo, Radicalização Coletiva, Multidões

Resumo alargado

Introduction

Neste capítulo são explicados os objetivos do trabalho, a motivação que nos levou a iniciar esta investigação, a abordagem que escolhemos para resolver o problema principal, o contributo que esta investigação traz à comunidade científica e, por último, é também apresentada a organização do documento para facilitar a compreensão do trabalho desenvolvido.
O início dos serviços de microblogging teve, nos último anos, um grande impacto na forma como as pessoas pensam, interagem, comportam, aprendem e fazem as suas atividades.
Quando um grupo de tweets, focado num tópico, que está a obter especial atenção por parte dos utilizadores, o mesmo pode constituir uma potencial ameaça. De acordo Krumm, [Kru13] é possível identificar o uso de uma linguagem radicalizada específica nas multidões de protestos, sendo aqui que reside o principal foco do da nossa investigação. Contudo, é importante reconhecer que a esmagadora maioria das informações postadas é inofensiva e representa multidões casuais, convencionais ou expressivas, além de ruído [BNG+11].
Devido a vários fatores, tais como a conveniência de utilização e a falta de regulamentação, a enorme quantidade de conteúdo gerado pelos utilizadores reflete mais de perto o mundo offline do que uma fonte de notícias oficial. Portanto, os social media, em específico as redes sociais, tornaram-se uma plataforma atraente para quem procura informação independente e, eventualmente, notícias mais realistas.
Recentemente, assistimos às notícias sobre os ”coletes amarelos” ou, em francês, ”Gilets Jaunes”. Começou por ser uma manifestação pacífica, mas alguns grupos extremistas infiltraram-se nas manifestações, tornando-a um protesto cada vez mais violento, como pudemos ver nas notícias. ”Ausência do progresso do movimento, da inexperiência dos manifestantes, da ação de grupos extremistas, das forças de altos deveres”, [Cre18].
Mais recentemente, em Portugal, assistimos a alguns eventos que podem ser considerados radicais. Enquanto algumas pessoas protestavam contra as ações da polícia num bairro problemático conhecido como o Bairro da Jamaica, outras pertencentes a um grupo extremista, protestavam de forma violenta contra os políticos que defendiam as pessoas desde o primeiro protesto [Rei19].

Definitions

Neste capítulo, são estudados alguns conceitos e definições relacionados com o extremismo, a radicalização e a radicalização coletiva. De seguida, há uma definição de redes sociais, explicando um pouco mais sobre o Twitter, são também apresentadas algumas definições sobre a área que abordaremos no tópico de social media. Para além disso, mencionamos várias definições resultantes da junção dessas áreas: extremismo, radicalização, radicalização coletiva e social medias.
É importante definir vários conceitos que usaremos neste documento e com essas definições podemos entender que este é um tópico muito sensível. Os conceitos que iremos definir são:

- Extremismo;
- Radicalização Coletiva;
- Multidão;
- Social Media;
- Análise de Sentimentos/Emoções/Opinião

State of the Art

O objetivo desta seção é criar um estado da arte para o projeto que estamos a desenvolver, que visa compreender o extremismo e a radicalização coletiva usando métodos que podem ajudar a obter um sistema não supervisionado e independente do idioma. Para isso, estudamos alguns projetos que têm semelhanças com o que estamos desenvolvendo.
Para satisfazer o objetivo de um sistema independente de idioma, confiaremos numa abordagem probabilística que pode ser aplicada em qualquer idioma ou mesmo numa mistura de idiomas.
Na parte não supervisionada, pretendemos criar um sistema que possa detetar tweets extremistas ou radicais por si só, o mínimo que pretendemos é que o sistema só precise ser treinado no início e faça o trabalho sozinho sem intervenção humana.
Existem vários métodos que, até agora, foram aplicados numa abordagem supervisionada, mas o método não está limitado a ele, podemos usá-lo e dar-lhe um caminho diferente. Como exemplo disso, temos a abordagem usada e descrita no artigo "Learning Sentiment-Specific Word Embedding for Twitter" [TWY+14]. A abordagem usada neste artigo é uma abordagem supervisionada, que não é o que pretendemos fazer, mas este artigo tem um método realmente interessante usado para um final diferente, mas com uma boa variedade de reutilização em outros objetivos.
É importante que nas análises de sentimentos sejam estudados diferentes níveis de extremismo e radicalização, pois é preciso ter cuidado com as pessoas definidas como extremistas ou envolvidas na radicalização coletiva. Um usuário pode estar a falar sobre um tópico que representa extremismo, mas não é extremista.
Como exemplo, sabe-se quando o ISIS começa a tornar-se mais ativo nas redes sociais, algumas contas que nada tinham a ver com grupos extremistas, como a conta de Barack Obama, foram temporariamente excluídas do Twitter. Portanto, deve-se evitar esse tipo de situação.

Extremism Understanding

Redes sociais on-line, como Facebook [1], Twitter [2] e Tumblr [3], tornaram-se uma plataforma de fato para centenas de milhões de usuários da Internet. As populares plataformas sociais, como Twitter, Facebook, Tumblr e muitas outras, são novos tipos de blog que facilitam a comunicação com pessoas de todo o mundo.
Numa extremidade, as postagens ou mensagens de tweet permitem que as pessoas entendam visões construtivas, ideias ou compartilhem artigos, vídeos ou links, etc.
Neste estudo, apresentamos uma abordagem inovadora para detetar sentimentos extremos nas medias sociais. O nosso objetivo é analisar corpus para identificar posts extremistas e usar o ExtremeSentiLex com polaridade negativa e positiva. Nós projetámos e desenvolvemos um aplicativo de protótipo chamado *Extreme Sentiment Generator* aplicando em dois recursos lexicais, SENTIWORDNET 3.0 [BES10] e SenticNet 5 [CPHK18], a fim de obter um recurso lexical extremo que contém termos positivos e negativos, o que representa sentimentos extremos usando métodos estatísticos, conforme descrito nas seções abaixo.
Além disso, também analisamos a compilação de conjuntos de dados para garantir a precisão da abordagem para encontrar sentimentos extremos nas postagens das medias sociais. Classificamos cinco corpus já utilizados e estudados pela comunidade científica, todos constituídos por posts encontrados nas medias sociais e alguns apenas com posts nas redes sociais. Utilizamos

[1] https://facebook.com/
[2] https://twitter.com/
[3] https: // www .tumblr.com /

uma matriz confusão para calcular o recall, a precisão e a exatidão do desempenho de nossa aplicação.
Este capítulo está organizado da seguinte forma: a Seção 2 apresenta brevemente os antecedentes; A seção 3 discute o trabalho relacionado; A Seção 4 demonstra a nossa abordagem metodológica para recolher o ExtremeSentiLex de sentimentos extremos; A seção 5 mostra a configuração experimental; A seção 6 apresenta e discute os resultados; A seção 7 conclui o capítulo.

Tests and Validation

Neste capítulo, são apresentados os resultados de todas as nossas experiências feitas com o ExtremeSentiLex. Fizemos experiências em vários conjuntos de dados com conteúdos diferentes, mas principalmente com informações de redes sociais e medias sociais. Para isso, analisamos a sua polaridade nos extremos dos conjuntos de dados rt-polarity, T4SA, Sentiment140, Ansar1 e TurnToIslam.
Depois de mostrar os resultados obtidos nas tabelas e gráficos, fazemos uma análise dos resultados para explicar o nosso ponto de vista. Iremos gerar mais gráficos para comparar a nossa abordagem com e sem incorporação de palavras e explicaremos a análise. Uma discussão comparativa nesta seção é importante, pois só podemos saber que temos bons resultados quando temos um termo de comparação.
Estamos a usar dois tipos de conjuntos de dados: temos alguns onde podemos comparar a polaridade do sentimento com a polaridade extrema que detetamos e outros que têm maior probabilidade de ter menos valores inconclusivos, uma vez que são conjuntos de dados obtidos em fóruns de radicalização.

Conclusion and Future Work

Neste capítulo são tecidas as considerações finais relativamente a todo o processo de desenvolvimento deste trabalho de investigação. Para além disso, são também exploradas abordagens que podem levar a um aprimoramento não só do ExtremeSentiLex mas também do processo de deteção de sentimentos extremos em posts de redes sociais.

Abstract

Increasingly in social media, we find cases where groups are organized to protest against something, often in those groups, members with extremist ideologies are inserted. These cases are happing more often, groups are created for the organization of peaceful protests and someone starts a topic with an extremist language leading, sometimes, to a radicalisation of the group.
This research aims to create an approach that allows the detection of cases of extremism and collective radicalisation within social networks, this should be done in an unsupervised and independent of language way.
The methods used to achieve the intended objectives are the creation of a lexicon of extreme sentiment terms named ExtremeSentiLex and a classifier of extreme sentiment in which the input is the extreme sentiment terms and the social network post. For the development of these tools were used purely statistical natural language processing methods. To validate the ExtremeSentiLex it was applied using the extreme sentiment classifier, the input posts that are analysed are posts from a dataset already validated by the scientific community. For a comparative study, word embeddings are used to expand the first ExtremeSentiLex obtained and a test is also performed in which the ExtremeSentiLex is balanced and applied to a balanced polarity dataset.
The results obtained in this content level research that will be available to the scientific community are the ExtremeSentiLex and several datasets that were evaluated by us regarding the presence of extreme sentiment. At the level of tests performed when the ExtremeSentiLex was validated, the level of precision in finding extreme sentiment at the correct polarity was very high. When applying word embeddings the results dropped. Regarding the ExtremeSentiLex and balanced dataset, the results were very positive.
It has been concluded that our dataset is suitable for the application in detecting extreme sentiments in text. Furthermore, it was found that with the help of linguistic and psychological experts the ExtremeSentiLex could be improved. However, this investigation aimed to do so using purely statistical methods. This goal has been successfully achieved.

Keywords

Sentiment Analysis, Extreme Sentiment Analysis, Social Media, Social Networks, Natural Language Processing, Statistic Natural Language Processing, Extremism, Collective Radicalisation, Crowds

Contents

1 Introduction **1**

1.1 Context of Research 1

1.2 Problem Statement and Goals 2

1.3 Approach to Solve the Problem 3

1.4 Document Organization 3

2 Theoretical Background **5**

2.1 Introduction 5

2.2 Extremism and Collective Radicalisation 5

2.2.1 Extremism 5

2.2.2 Collective Radicalisation 6

2.2.3 Crowd 7

2.3 Social Media 7

2.4 Sentiment/Emotion/Opinion Analysis 9

2.4.1 Sentiment Analysis Applications 9

2.4.2 Sentiment Analysis on Social Networks 10

2.5 Conclusion 11

3 State of the Art **13**

3.1 Introduction 13

3.2 Related Works 13

3.2.1 The roots of radicalism 13

3.2.2 Terrorism Detection using SA and ML 15

3.2.3 Learning with Hashtags 18

3.2.4 How to be Unsupervised 21

3.2.5 How to be Language-Independent 21

3.2.6 Specific possible methods and tools to be used 22

3.3 Word Embeddings 24

3.4 Conclusion 25

4 Extremism Understanding **27**

4.1 ExtremeSentiLex 27

4.1.1 Used Lexical Resources 28

4.1.2 Defining Extreme Polarity 28

4.1.3 Generating ExtremeSentiLex 29

4.2 Detect Extreme Sentiments 30

4.2.1 Datasets 30

4.2.2 Experimental Setup 31

4.3 Conclusion 33

5 Tests and Validation **35**

5.1 Tests and Results 35

5.1.1 Results for ExtremeSentilex 36

5.1.2 Results for ExtremeSentilex with word embeddings 38

5.1.3 Results for Balanced Corpus and Dataset 39
5.2 Discussion 41
5.3 Conclusion 45

6 Conclusion and Future Work **47**
6.1 Conclusion 47
6.2 Future Work 48
6.2.1 Radicalisation 48
6.2.2 Human contributions for the improvement of ExtremeSentiLex 48
6.2.3 Machine Learning Techniques 49

Bibliography **51**

A Appendix **57**
A.1 ExtremeSentiLex Content 57

List of Figures

2.1 Social Media usage in 2019 8
2.2 Sentiment Analysis Applications over the Years 10

3.1 Sentiment Classification Techniques Used in SA [MHK14] 16
3.2 Proposed system development diagram [Isk17] 17
3.3 Diagram explaining that even with the same prefix it might not transmit the same emotion 19
3.4 Bootstrapped Learning Diagram used by Qadir and Riloff. (HT = hashtag; HP = hashtag pattern) [QR14] 20
3.5 Example of words vectors 25

4.1 Extreme sentiment collection process. 27
4.2 Process of testing of Extreme Sentiment Classifier 31

5.1 Accuracy, Recall and Precision Percentage per Dataset 37
5.2 Accuracy, Recall and Precision Percentage per Dataset for lexicon with word embeddings 39
5.3 Accuracy, Recall and Precision Percentage per version of ExtremeSentiLex applied to B-T4SA 41
5.4 Comparison between results of rt-polarity applying ExtremeSentiLex with and without word embeddings 42
5.5 Comparison between results of T4SA applying ExtremeSentiLex with and without word embeddings 43
5.6 Comparison between results of Sentiment140 Training applying ExtremeSentiLex with and without word embeddings 43
5.7 Comparison between results of Sentiment140 Test applying ExtremeSentiLex with and without word embeddings 44
5.8 Comparison between Inconclusive results in all datasets when applying ExtremeSentiLex with and without word embeddings 44
5.9 Comparison between results of T4SA applying ExtremeSentiLex with word embeddings and the balanced ExtremeSentiLex 45

List of Tables

5.1 rt-polarity results 36
5.2 T4SA results 36
5.3 Sentiment140 training results 36
5.4 Sentiment140 test results 36
5.5 TurnToIslam and Ansar1 results 36
5.6 Indicators of algorithm efficiency 37
5.7 rt-polarity results for lexicon with word embeddings 38
5.8 T4SA results for lexicon with word embeddings 38
5.9 Sentiment140 training results for lexicon with word embeddings 38
5.10 Sentiment140 test results for lexicon with word embeddings 38
5.11 TurnToIslam and Ansar1 results for lexicon with word embeddings 38
5.12 Indicators of algorithm efficiency for lexicon with word embeddings 39
5.13 Results of B-T4SA with ExtremeSentiLex 40
5.14 Results of B-T4SA with Extended ExtremeSentiLex 40
5.15 Results of B-T4SA with balanced ExtremeSentiLex 40
5.16 Indicators of algorithm efficiency for ExtremeSentiLex whit B-T4SA 40

Acronyms

API	Application Programming Interface
CF	Collaborative Filtering
ESC	Extreme Sentiment Classifier
ESG	Extreme Sentiment Generator
ICT	Information and Communication Technologies
ISIS	Islamic State of Iraq and Syria
JSON	JavaScript Object Notation
HTML	HyperText Markup Language
ML	Machine Learning
NLP	Natural Language Processing
NLTK	Natural Language Toolkit
OSN	Online Social Network
SA	Sentiment Analysis
SNLP	Statistical Natural Language Processing
URL	Uniform Resource Locator
XML	Extensible Markup Language

Chapter 1

Introduction

In this chapter will be explained what are the goals of the work, the motivation that leaded us to this research, the approach we choose to solve the main problem as well as the contribution this research gives to the scientific community and finally the organization of this document for an easier understanding of the developed work.

1.1 Context of Research

In the last few years, the advent of micro-blogging services has been impacting the way people think, communicate, behave, learn, and conduct their activities.
Due to a number of factors, including the convenience of use and the lack of regulation, the huge amounts of user-generated content reflects more closely the offline world than the official news source. Therefore, social media have become an attractive platform for anyone seeking independent information and eventually more realistic news.
We have recently been assisting the news about the "Yellow vests" or in French "Gilets Jaunes". It started as a pacific manifestation but some extremist groups have joined the manifestations, so it became a very violent protest as we can see from the news. "Absence of the progress of the movement, the inexperience of the demonstrators, the action of extremist groups, the forces of higher duties", [Cre18].
More recently, in Portugal, we have assisted to some events that can be taken as radical. While some people were protesting against the actions taken by the police on a problematic neighbourhood known as Bairro da Jamaica there were other from an extremist group protesting against the politicians, in a violent way, who defended the people from the first protest [Rei19].
Each cluster of social networks posts focusing on a bursty topic may constitute a potential threat. However, the majority of clusters are armless and represent casual, conventional or expressive crowds as well as noisy data [BNG+11]. To identify acting or protest crowds, we propose to understand the typical language usage present in each cluster as well as its network activity. Indeed, ultimately, a crowd is characterized by its dominant emotion, its level of interaction and shared focus.
Krumm[Kru13] showed that specific radicalised language is used within acting and protest crowds. Therefore, we propose that each tweet inside a cluster is classified as radical or non-radical in terms of language use, so that the collective radicalisation of a cluster can be measured. As far as we know, there is no previous work on modelling radicalised language. Radicalisation is a process by which an individual or group comes to adopt increasingly extreme political, social or religious ideals and aspirations . As such, we hypothesize that radicalised language mainly expresses negative emotions (such as anger, fear, or anxiety) and also positive emotions (such as joy, love, or devotion) with high intensity, following the classification of Plutchik wheel of emotion [PK80].
There are a great deal of works aiming at classifying tweets in some intended categories. In all cases, the main problem lays in the fact that only a few training examples can be afforded,

and so robust classifiers cannot be built. Traditional methods to solve this problem include self-taught, semi-supervised or ensemble learning. In this specific context, we propose a solution based on self-taught learning [RBL+07]. For that, we aim to use methods based on statistical NLP.
Self-taught learning aims at defining higher-level feature representations based on unlabelled data that can easily be gathered. Word embeddings [MSC+13] classically propose such high-level representations. Some successful results have recently been obtained in the domain of sentiment tweet classification [TWY+14], upon which we propose to improve. Although word embeddings have shown great improvements over most NLP tasks, they show some limitations. As most models exclusively use syntactic contexts [MSC+13], words with different connotations may not be well-separated in the space, and consequently produce error-prone classifiers. Emotional words, such as fear and joy are an example.

1.2 Problem Statement and Goals

This thesis proposes to learn weak classifiers based on dictionaries of emotional words [QR14] to retrieve many roughly classified emotional tweets, in a similar way as [TWY+14]. The continuous high-dimensional space can then be learned by including both syntactic and emotional contexts into a recurrent neural network. However, only highly intensive emotional language may identify threatening crowds and not just emotional language. Intensive language can be seen as the difference between ”angry” and ”wild”. Both words share common semantics but with different intensity levels. In a well-behaved semantic space, angry and wild should also be separated, which is not the case by only looking at the syntactic and emotional contexts. Therefore, we propose to build word embeddings that consider emotional and intensity contexts together with syntactic context. For that purpose, weak classifiers for intensity detection will be built based on recent work of Sharma et al. [SGAB15] on language intensity. It is interesting to notice that we will build a continuous semantic space, where both lexical items and named entities are joined. As such, entities such as Adolf Hitler and Winston Churchill may clearly be separated in such a semantic space.
Finally, this thesis will study the introduction of demographically-driven word embeddings. Recently, Bamman et al. [BDS14] proposed to develop word embeddings considering the localization of the issuer of the conveyed message. These findings open a great deal of improvements, as tweets can be geo-localized but also include information such as age and gender of the issuer. We deeply believe that localization and age can drastically improve the correct encoding of words in continuous spaces and, therefore, produce high-performing classifiers for radicalised language. Indeed, radicalism may not be expressed in the same way whether the issuer of a tweet is 18 or 45 years-old. Moreover, geo-localization can consider specific regional language usage. As far as we know, there are no studies on demographically-driven word embeddings. This thesis proposes to build word embeddings in a similar way as Bamman et al. [BDS14], but including both age and gender as well as geo-localization and all its combinations.
The major problem we aim to solve is the detection of extremism and radicalisation generated online, specially on social networks, such as Twitter.
As a conclusion, we aim to avoid crowds which may turn into violent crowds if they are born from online platforms where the motivations of this crowds are based on extremism and/or radicalisation. We will focus our research on social media, such as social networks (e.g. Twitter), forums and website comments.

1.3 Approach to Solve the Problem

We will apply SNLP to solve the problem presented above. We will start by studying the previous works developed in the area, specially works involving extremism and radicalisation, and detection of extreme sentiments on texts from social media.
Once we have understand the state-of-the-art, we will start developing an approach to solve the problem. To start, we will create an extreme lexical resource which will be a compilation of terms with extreme intensity from two lexical resources already validated by the scientific community. After the generation of our extreme sentiment lexicon, we will apply it to datasets created for the study of the language, specially in computer science, in order to validate our lexicon.
For the generation of the extreme sentiment lexicon we will need to create measures and algorithms to identify the terms considered extremes. Also, we will need to evaluate the text samples on the datasets and detect the existence of extremism or not.
We hope our work helps others to invest in the areas once this is a nowadays problem. We will not left our work just for us. We contribute with many files which may be used by the community for further investigations on the area of detecting and fight extremism and radicalisation, specially on online platforms. Also, in this thesis will be written a state-of-the-art to help us in the develop of our work, and also other investigators on building better approaches using futures methodologies. We will provide versions of our extreme sentiment lexical resource which we called ExtremeSentiLex, one version with only the terms extracted from the both Senticnet and SentiWordNet, the second version with the terms we added from word embeddings. Furthermore, we also provide the datasets with our analysis identifying posts as extreme positives, negatives or inconclusive, as we cannot tell if they are extreme or not.

1.4 Document Organization

This thesis is organized in six chapters:

- The first chapter is the introduction where we find the context of this research, the goals of our work and statement. After this, we lightly explained the approach we will apply to solve our problem. We show our contribution to the scientific community and finally the organization of the document;
- In the second chapter we find the definitions we need for the development of this thesis and all the work.It is important to define several concepts we will use in this paper and with those definitions we can understand that this is a really sensitive topic. The concepts we will define are extremism, collective radicalisation, crowd, social media, sentiment/emotion/opinion analysis;
- The third chapter is the state-of-the-art where we made a research on the the area, study and learned about the relevant work developed so far by other authors. The purpose of this chapter is to create a state-of-the-art for the project we are developing, that aims to Understand the Extremism and Collective Radicalisation by using methods that can help to obtain a unsupervised and language independent system. In order to do that, we have been studying some projects that have similarities with the one we are developing;
- In the fourth chapter we will deeply explain our approach and we also write about the dataset we used. In this study, we present an innovative approach to detect extreme

sentiments on social media. We aim to analyse corpora in order to identify extremist posts and used ExtremeSentiLex having a negative and positive polarity. We also analyse the compilation of datasets in order to ensure the accuracy of the approach to find extreme sentiments on social media posts.;

- Fifth chapter is where we show our tests and the results of the application of this tests. In the end we write a discussion where we analyse our results and take our conclusions. In this chapter are showed the results of all our experiments made using ExtremeSentiLex. We made experiments on various datasets with different contents, but mostly with information from social networks and social media. For that, we analysed its polarity on the extremes the datasets rt-polarity, T4SA, Sentiment140, Ansar1, and TurnToIslam. After showing the results obtained on tables and graphics we make an analysis on results to explain our point of view. We will generate more graphics to compare our approach with and without word embeddings and explain it making an analysis.;

- For the last chapter we have the conclusions from the entire thesis and also the future work that can be done in the future. In this chapter the final considerations are made regarding the whole process of development of this research work. In addition, approaches are explored that may lead to an improvement not only of the lexicon but also of the process of detecting extreme feelings in social network posts;

- Finally we have the bibliography where we have our references.

Chapter 2

Theoretical Background

2.1 Introduction

In this chapter some concepts and definitions related to the Extremism, Radicalisation, Collective Radicalisation are studied. After that, we will define social networks, explaining a little bit more about Twitter, as well as introduce some definitions on the area that we will approach in the topic of social media. Furthermore, several definitions resulting from joining these areas Extremism, Radicalisation, Collective Radicalisation and social media.
It is important to define several concepts we will use in this paper and with those definitions we can understand that this is a really sensitive topic. The concepts we will define are:

- Extremism;
- Collective Radicalisation;
- Crowd;
- Social Media;
- Sentiment/Emotion/Opinion Analysis

2.2 Extremism and Collective Radicalisation

This section will have the definitions of Extremism, Radicalisation, and Collective Radicalisation for better understanding the area and what we want to study with this thesis.

2.2.1 Extremism

Scruton [Scr07] defined extremism as a vague term which can be explained by three meanings:

1. "Taking a political idea to its limits, regardless of 'unfortunate' repercussions, impracticalities, arguments, and feelings to the contrary, and with the intention not only to confront but also to eliminate, opposition".

2. "Intolerance towards all views other than individual own".

3. "Adoption of means to political ends which disregard accepted standards of conduct, in particular which show disregard for the life, liberty and human rights of other."

As we understand from Lowe article[Low17], there are two types of extremism the non-violent and the violent, the difference between them is that the first one is the most common one, and sometimes it gets undetected. One example of non-violent extremism are posts on Twitter about an idea that appeals against all the other ideas. On the other side, there is violent extremism. As an example we have the worldwide riots, as there is on Chile[Mer19]. There are many cases nowadays where we can see the two ways of extremism, where we can see extreme ideals getting into society.

2.2.1.1 Extremism on Social Networks

Many users are using social networks for various ends. Unfortunately, some use it to spread distorted beliefs, negative influence on other users, like terrorism, extremism, and radicalism ([Isk17]). Since mid-2015 Twitter has already deleted more than 125,000 accounts that were some way linked to terrorism ([Yad16]). Focusing on social networks the definition of "extremist" can vary. In a social network, extremist users can exhibit different posting patterns, which must be recognized. For instance, one can have a long history of moderate yet radical posts, slowly disseminating his view and doctrines, whereas another one can have quite shorter posting history but much more intense, in terms of content and level of extremism, implying that the latter is already a radicalised figure, while the former is in the process of becoming ([SDF18]). As mentioned above, we can conclude that due to the easy and free communication which is offered in social networks, people easily receive and watch information related with extremisms. Nowadays people easily get extremist from simple ideas which can be an issue. We need to stay tuned to people, not just in their daily lives, but also on social media where people can have a different personality. Sometimes people take posts on social networks so serious and sometimes have an extreme vision of things, that it is transported to the quotidian in an extreme and problematic way.

2.2.2 Collective Radicalisation

Defining radicalisation is known to imply a movement in the direction of supporting or enacting radical behaviour ([KGB+14]). We may consider a radical behaviour when serving a given end, it undermines other goals that matter to most people ([KGB+14]). As mentioned above, it can be concluded that collective radicalisation is the effect of it but on a group, creating crowds on roads showing their point of view and sometimes becoming violent. In the next section we will define what a crowd is. As we understand from the work of Schmid[Sch13], there is no universal concept for radicalisation in academia or government. The concept of radicalisation is by no means as solid and clear as many seem to take for granted. The Expert Group on Violent Radicalisation established by the European Commission in 2006, tasked to analyse the state of academic research on radicalisation to violence, in particular terrorism. In 2008 noted that '[r]adicalisation is a context-bound phenomenon par excellence. Global, sociological and political drivers matter as much as ideological and psychological ones'[ABDP+08].From this group we got a definition of violent radicalisation, 'socialization to extremism which manifests itself in terrorism'.[ABDP+08]

In the special issue of the International Journal of Conflict and Violence (2011)[DPL12], the guest editors Donatella della Porta and Gary LaFree, quoted many definitions and we understand that all of them are suitable:

- [...] in the 1970s, the term radicalisation emerged to stress the interactive (social movement/state) and processual (gradual escalation) dynamics in the formation of violent, often clandestine groups (Della Porta, 1995). In this approach, radicalisation referred to the actual use of violence, with escalation in terms of forms and intensity;
- Radicalisation may be understood as a process leading towards the increased use of political violence[...];
- [...]radicalisation is understood as an escalation process leading to violence;

- Many researchers conceptualize radicalisation as a process characterized by increased commitment to and use of violent means and strategies in political conflicts. Radicalisation from this point of view entails a change in perceptions towards polarizing and absolute definitions of a given situation, and the articulation of increasingly 'radical' aims and objectives. It may evolve from enmity towards certain social groups, or societal institutions and structure. It may also entail the increasing use of violent means.

- It may be more profitable to analyse radicalisation as a process of interaction between violent groups and their environment, or an effect of interactions between mutually hostile actors;

- Functionally, political radicalisation is increased preparation for and commitment to intergroup conflict. Descriptively, radicalisation means change in beliefs, feelings, and behaviours in directions that increasingly justify intergroup violence and demand sacrifice in defence of the group;

- Radicalisation [...] can be understood to be the strategic use of physical force to influence several audiences.

2.2.3 Crowd

A crowd refers to a temporary gathering of people united by a common focus, where individuals are known to influence each other. Formally, crowds can be divided into five distinct categories: casual crowd, conventional crowd, expressive crowd, acting crowd and protest crowd ([RS06]).
A crowd may happen in physical space, in a digital platform or it can be hybrid. As an hybrid example we have the "yellow vests" event. It was initialized in a digital platform and the results appeared as protests on the streets of France.
The crowds we want to study are the digitals once our focus is on social medias. This kind of crowd rises from specific material-discursive arrangements [AB13]. "Instead of appearing only in local hotspots, ICT enable crowds to transcend traditional geographical limits. Technology supports the quick formation of loose social ties across a geographically dispersed population"[JDC19]. Another important aspect in many of these crowd instantiations is their globalizing of resources. Thanks to social media, crowd networks do not only quickly expand beyond local contexts, but, through their geographical transcendence, may also globalize the influx of resources. Uprisings profit from logistical, material or informational support from 'outsiders'[Lim12].In social protest and uprisings, networks extend beyond local confines and even beyond national boundaries [SP12].
As we saw on the protests of "yellow vests" it started on France and was organized on a social network. But due to the social media as rapidly spread for more countries specially on Europe. And it was impossible to prevent due to the lack of power from the authorities in this platforms.

2.3 Social Media

According to [Shi08], social media is a tool that "increases our ability to share, to cooperate with one another, and to take collective action, all outside the framework of traditional institutions and organizations". So it can be seen that social media is a good tool to urge crowds on one solid focus, once people cooperate and share information this can associate people from different places with the same objective.

The comprehension and definition of social media requires a clear understanding of society, including what does it mean to be social and what exactly is a social act. The best field to answer these questions is *Social Theory*, a sub-field of *Sociology* ([Fuc17]).
The use of social media and messaging applications grew 203% year-on-year in 2013 as reported by Statistician, quoting data from Flurry Analytics, as we can see in Figure 1. This growth means that 1.61 billion people are now active in social media around the world and this is expected to advance to 2 billion users in 2016, led by India.

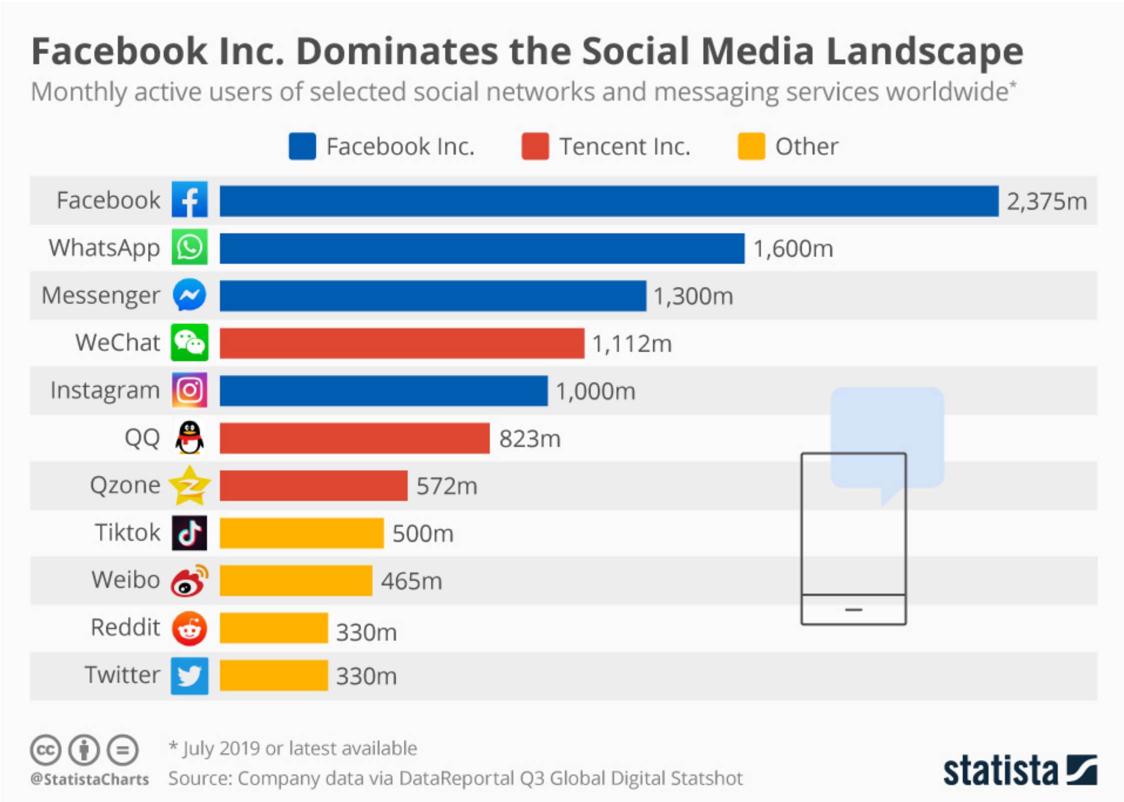

Figure 2.1: Social Media usage in 2019
Source: Statista (http://bit.ly/2Q1Mj3t)

However, the huge amount of daily posts do not allow an accurate assimilation of information by human users ([Dia10]). This problem is known as information overload. As a consequence, automatic services capable of understand, classify, summarize and verify information are increasingly needed by human users as well as public authorities and private companies. This situation has given birth to a new research area called urban informatics, that aims to exploit the large quantities of information produced by modern cities in order to gain insights into how they function. These insights lay the foundation for improving the lives of citizens, by improving the efficiency of public services, and satisfying complex arising information needs ([Fot08]).
Social media is not easily defined. But we can understand that it came with the Web 2.0 ([Fuc17]), and it can be a blog, a website or even a social network such as Facebook [1] or Twitter [2]. Nowadays, social networks are one of the most powerful weapons on changing public opinion. People use these networks to organize mobs and riots, for example, and also it is very easy to recruit people with extremist ideologies for extremist groups ([NEW17]). We see in newspapers that extremist groups are on Twitter reading people tweets to identify potential

[1] https://facebook.com/
[2] https://twitter.com/

members. ISIS publishes a monthly online magazine to make people travel to Syria using social media ([Obs15]).

2.4 Sentiment/Emotion/Opinion Analysis

”Sentiment Analysis or Opinion Mining is the computational study of people’s opinions, attitudes and emotions toward an entity. The entity can represent individuals, events or topics” ([MHK14]). Opinion Mining is not the same as Sentiment Analysis (SA). Opinion Mining starts by extracting and analysing the opinion about something while SA is more about the sentiment that something causes on people, usually expressed on text, like in tweets or Facebook posts ([MHK14]). Others say that SA can be seen as a type of text classification but deals with subjective statements that are harder to classify ([KDA$^+$16]).
SA uses NLP algorithms to analyse the sentiment on words we get by using them. SA is the set of computational techniques to extract, classify, understand and assess the opinions expressed in online sources, such as social media comments, and other user-generated contents ([CH12]).
SA has been used on many areas to improve results. For example, we can use it to know how the market is behaving and, with that, we can predict what is happening next on that market. For this theme, we are using it to identify extreme behaviour on social media and more specifically on social networks because this is our target: find people who represent a risk as extremist and identify groups, detecting if there exists the possibility of creating a group of radical people or, in other words, collective radicalisation.
SA may be considered a classification process. It can be divided into three levels: document-level, sentence-level, and aspect-level. In document-level, SA classifies an opinion document about its polarity (negative or positive). It should take into account the entire document as a basic information unit. In sentence-level, SA is the expressed sentiment that is classified in each sentence. We first identify if the sentence is subjective or objective. And if it is subjective sentence-level SA will determinate the polarity of the opinion. Using these two levels does not provide the necessary detail. To obtain more detailed SA on the opinion we need to get to the aspect-level. Aspect-level SA, classifies having in count the specific aspects of entities. First, we identify the entities and their aspects, it can give different opinions for different aspects of the same entity ([MHK14]).

2.4.1 Sentiment Analysis Applications

Nowadays, SA has numerous applications, specially because of the social networks. There are already some areas that would not survive without SA because they had to use this tool in order to follow the evolution of their users/clients. Next, a brief description of the areas that use SA in order to get better results and how they use in the following:

1. Marketing: Since social media has become a unique platform of customer interactions, the use of SA can easily take marketing to a whole new level. Companies have figured that emotions of social media are shaping their brands image. Therefore, SA tools give marketeers a way to measure their effectiveness, and help consumers who are trying to research a product or a service ([DMSR14]).

2. Politics: Many valuable uses can be obtained for political organizations by fully understanding social media sentiment. Social media feedback has been used to inform political

leaders of potential threats, problems or issues with their organizations. In addition, an essential role for SA appeared earlier in predicting elections, and acquiring citizens responses on important issues such as increasing prices and changing the constitution ([DMSR14]).

3. Health care: Medical web blogs are all over the internet these days. These web blogs contain only medicine and health-care issues such as diseases, medical treatments and medications. Due to the health-related experiences and medical histories these web pages provide for practitioners and patients, SA tools had to be developed for the use in medical fields ([DMSR14]).

4. Finance: SA can also be used in the financial world. Investors can easily follow their favourite assets and monitor their sentiment data in real time. With SA, business investors can acquire business news easier and aggregate this information to make better financial decisions ([DMSR14]).

The field of SA is becoming more important everyday. Institutions using new SA features in their systems are becoming more competitive, seeing huge growth in their revenues, as can be seen in Figure 2. SA can be important as well for the consumer, in many different business contexts. Systems equipped with SA features can recommend better choices, based on sentiment crowd sensing, previously acquired.

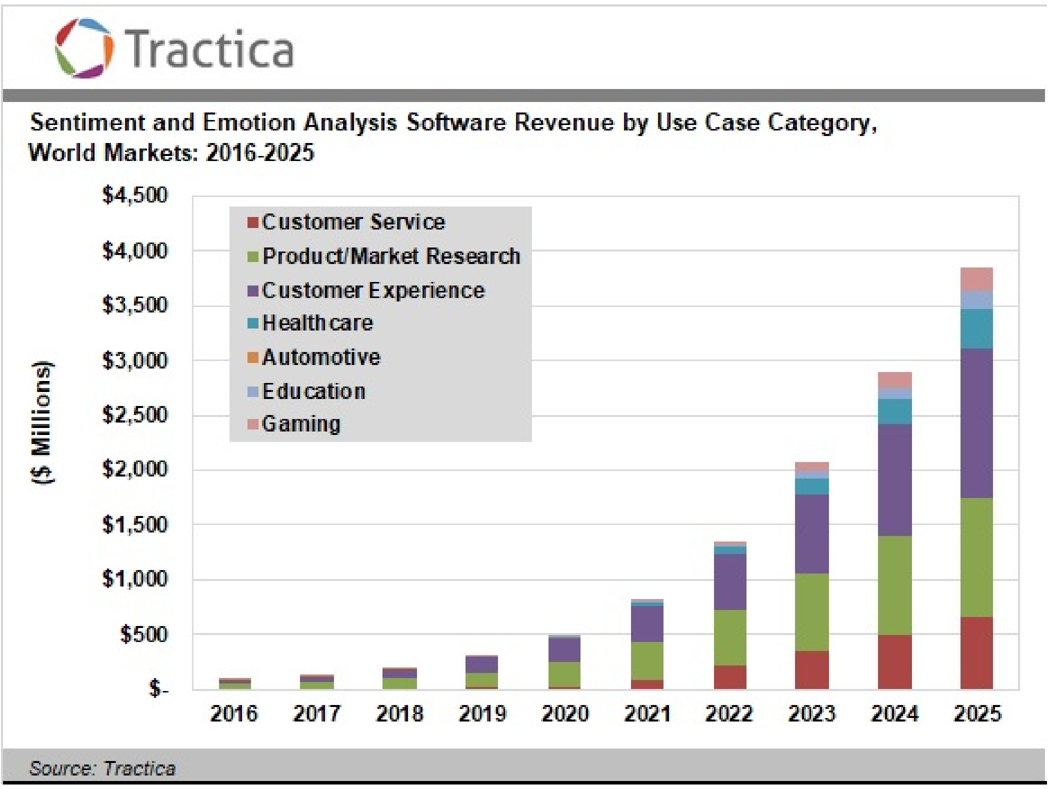

Figure 2.2: Sentiment Analysis Applications over the Years
Source: Tractica (https://bit.ly/2B1qCMY)

2.4.2 Sentiment Analysis on Social Networks

The main focus in this survey is exclusively on SA related to extremism and radicalism. And for that reason we will focus on SA on Social Networks like Facebook, Twitter, Instagram, and many others. Furthermore, as it focus on text, it makes sense that we explore more about Twitter,

that is based on text posts, unlike Instagram that is focused on images or Facebook, where we can find videos, images or text to convey feelings.
When we analyse social networks we are analysing structured and unstructured data ([GH15]). We can get this data by using the API's of the specific social network.Structures data represents only 5% ([Cuk10]) of all existing data, and it refers to the tabular data found in spreadsheets or relational databases. On the other hand, there is unstructured data like text, images, audio and videos, which sometimes lacks the structural organization required by machines for analysis ([GH15]). There is also semi-structured data that is data we can not categorize as structured or unstructured, XML, textual language for exchanging data on the web are some examples of semi-structured data. XML documents contain user-defined data tags which make them machine-readable ([GH15]).
We will analyse social networks to detect extremism and radicalism, find groups of extremist, communities and potential extremists. Our work relies on understanding extremism and collective radicalisation using an unsupervised and language independent approach. To do that, we will analyse and extract information from the posts on a social network. And using the extracted information the main focus is to avoid big crowds that can create trouble for all citizens and violence on the streets.
Using social networks, SA will help us discover this crowds even before they get to the streets and let the authorities take control of the situation or at least be better prepared for cases like the "yellow vests". "Yellow vests" started to organize on social networks and were incited by posts on social networks especially by Twitter where we can already find some hashtags related to this protest that as became violent and dangerous for all citizens. This inorganic movement started as a fight from the people to claim their rights, but in the middle of a pacific movement that try to strike the governments without brute force, there are some extremist groups that infiltrate on the crowds in order to create problems with the authorities in order to denigrate the image of the government in work.
We know that identifying emotions in the social networks can "help companies understand how people feel about their products, or to assist governments in recognize growing anger or fear associated with an event, or to help media outlets understand people emotional response toward controversial issues or international affairs" ([QR14]).

2.5 Conclusion

With this section, our objective was to obtain as much information as possible in different areas. Once our target is multidisciplinary it makes sense to obtain information for all the areas involved in the research.
We know social media have a big rule in the process of extreme ideas dissemination and all over the world Twitter is the main platform used to spread hate speech and extreme ideas. When we have dissemination of similar information from many users we might be witnessing a collective radicalisation.
We are not only working to prevent this kind of event but also an event that could mean harm to the citizens, tourists and etc. We are trying to predict an event like mobs, riot and so on. As we saw with the case of yellows vests, this inorganic social movement began on social networks especially on Facebook and Twitter by the use of groups, events, and hashtags.

Chapter 3

State of the Art

3.1 Introduction

The purpose of this chapter is to create a state-of-the-art for the project we are developing, that aims to Understand the Extremism and Collective Radicalisation by using methods that can help to obtain a unsupervised and language independent system. In order to do that, we have been studying some projects that have similarities with the one we are developing.
There are several methods that, until now, have been applied in a supervised approach but the method is not limited to it, we can use it and give it a different path. As an example of that we have the approach used and described on the paper ”Learning Sentiment-Specific Word Embedding for Twitter” [TWY+14]. The approach used on this paper is a supervised approach, that is not what we aim to do, but this paper has a really interesting method used for a different ending but with a good range of re-utilization in other objectives.
It is important that in sentiment analyses it is studied different levels of extremism and radicalisation, because it needs to be careful on the people defined as extremist or that are involved on collective radicalisation. A user can be talking about a topic that represents extremism, but not being extremist.
As an example, it is known when ISIS start to become more active on social networks, some accounts that had nothing to do with extremism groups, like the Barack Obama account, were temporally deleted from Twitter. So it must be avoided this kind of situations.

3.2 Related Works

This research has shed some light on the related works with an in-depth study for three works to an approach to the extremism and collective radicalisation.The following subsection describes related works involved in this paper.

3.2.1 The roots of radicalism

According to Fernandez et Al. [FAA18], there is an innovative approach to the topic of radicalisation and social networks, based on NLP and Collaborative Filtering (CF). The different roots of radicalisation are captured (micro-roots, meso-roots and macro-roots [Sch13], and each user is represented through keyword-based vector description.
They have also incorporated a theoretical perspective into a computational approach by doing what was written above. That has helped the authors develop an effective radicalism detection and prediction approaches. Different roots of radicalism were used in this work: micro-root, meso-root and macro-root. As we know, when a user participates in a social network they can take to ways on posting something:

1. Create and posting new content;

2. Sharing content posted by someone else.

In the work, they assumed that micro-roots are captured by all the content that the user has created. Similarly, meso-roots come from the shared posts by the user. And we have macro-roots that are captured by all the content that are external to the given social network (links/URLs), from other website or other social networks, videos, etc [FAA18].
To apply this roots to study radicalisation, they used vectors that include the posts of the users that represent each one of the roots. The vector representing micro-roots and meso-roots influences over a user were transformed into n-grams(uni-grams, bi-grams, and tri-grams) [FAA18]. Then, the value of each n-gram in micro-root vector of a user is computed as the frequency of the n-gram in the posts created by the user, normalized by the number of posts created by the user [FAA18]. The value of each n-gram in the meso-roots vector of a user is computed as the frequency of the n-gram in the posts shared by the user, normalized by the number of posts shared by the user [FAA18].
In the case of macro-roots influences, an automatic data scrapping over the URLs included on a macro-roots vector were performed, by automatically parsing the HTML and extracting the title and description of the websites. Giving the set of n-grams obtained after pre-processing all the links defined the macro of the user. The value of each word in a macro-roots vector of the user is computed as the frequency of the n-gram in all URL entries shared by the user, normalized the number of URL [FAA18].
They collected, integrated and extended existing lexicons. The used lexicons were:

- ICT Glossary;
- Saffron Experts;
- Saffron Dabiq Magazines;
- Rowe and Saif.

The merge of these lexicons was made to consider them as one unique lexical entry the term and their variances. They first incorporated syntactic variances of each term, particularly:

- Lowercase;
- Removal of apostrophes;
- Removal of hyphens;
- Removal of diacritics.

In other words if one lexicon contain one term that exists in other lexicon, they are merged into one unique entry in the final lexicon. They obtained a final lexicon with 305 entries, including 556 terms, expressions and variances. In order to compute the radicalisation influence of different roots over the user, they have compute the cosine similarity between the micro-roots and the meso-roots vectors and the generated lexicon. They were not able to compute the macro vectors due to a lot of site were already closed, and they could not gather the information from URLs.
CF strategies were used in this work to make an automatic prediction about the interests of a user by collecting preference information from many users [SLH14]. This approach can only be used by doing two steps:

1. look for users that have a similar rating pattern to the users for whom the prediction is made;
2. use the ratings of user found in the previous step to compute the predictions for the active user.

In their model, items are n-grams and ratings are the values of those n-grams in the posts created and shared by the users on social networks. The purpose for the usage of CF was to predict the future micro-roots, meso-roots and macro-roots influences for user.

They used two publicly available datasets, from Kaggle data science community, made to study radicalisation. One of the datasets includes 17,350 tweets from 112 pro-ISIS accounts. The second dataset was created as an opposite of the previous one. It contains 122,000 tweets from 95,725 users collected in different days.

They concluded that "creating intelligent technologies to automatically identify online radicalisation is a key priority of counter-extremist agencies. However, little effort has been devoted to integrate the knowledge of existing theories of radicalisation in the development of these technologies. In this paper we propose a computational approach for detecting and predicting the radicalisation influence a user is exposed to, grounded on the concept of 'roots of radicalisation', identified in social science models. While our approach constitutes a first step to bridge these disciplines, a stronger collaboration is needed to effectively target the problem online radicalisation" [FAA18].

It is important that radicalisation has a multi-pronged approach, based on the areas that can be used to combat the radicalisation effectively, to have more effective work. Is very important to detect and create different approaches to this topic so we can finally obtain the most effective approach using a mix of the best ideas used along the way.

3.2.2 Terrorism Detection using SA and ML

According to [Isk17], there is a new approach to a similar problem like the one on the above section. Knowing that social media has recently been the most important channel for people to interact and share ideas with each other. People chose to express their opinion about everything on social media because it is easier and faster to share online than use a complain book. Furthermore, online reaches more people and influences the choices of the potential users about the topic people are talking about.

Besides that, an even worst use of social media and networks, the share of extremist ideas and the recruit of members for extremist groups that want to be known all over the globe. In 2015 more the 250,000 accounts were linked to terrorism having that in mind the accounts have been deleted [Yad16].

As we already know the used methods are not the better ones. The current SA technique used to find if a tweet can lead to an extremist user is not really accurate. Nowadays, it is normal for a humorists to have an account on social networks, especially on Twitter, so for the technique used is hard to know if a tweet is a joke or a real threat, and we know that humorist makes jokes with everything, even with terrorism.

As shown in figure 3 two possible approaches for SA: lexicon-based approach and machine learning approach. These two approaches have been studied, [Isk17] decided to use the machine learning approach, but both of them have advantages and disadvantages.

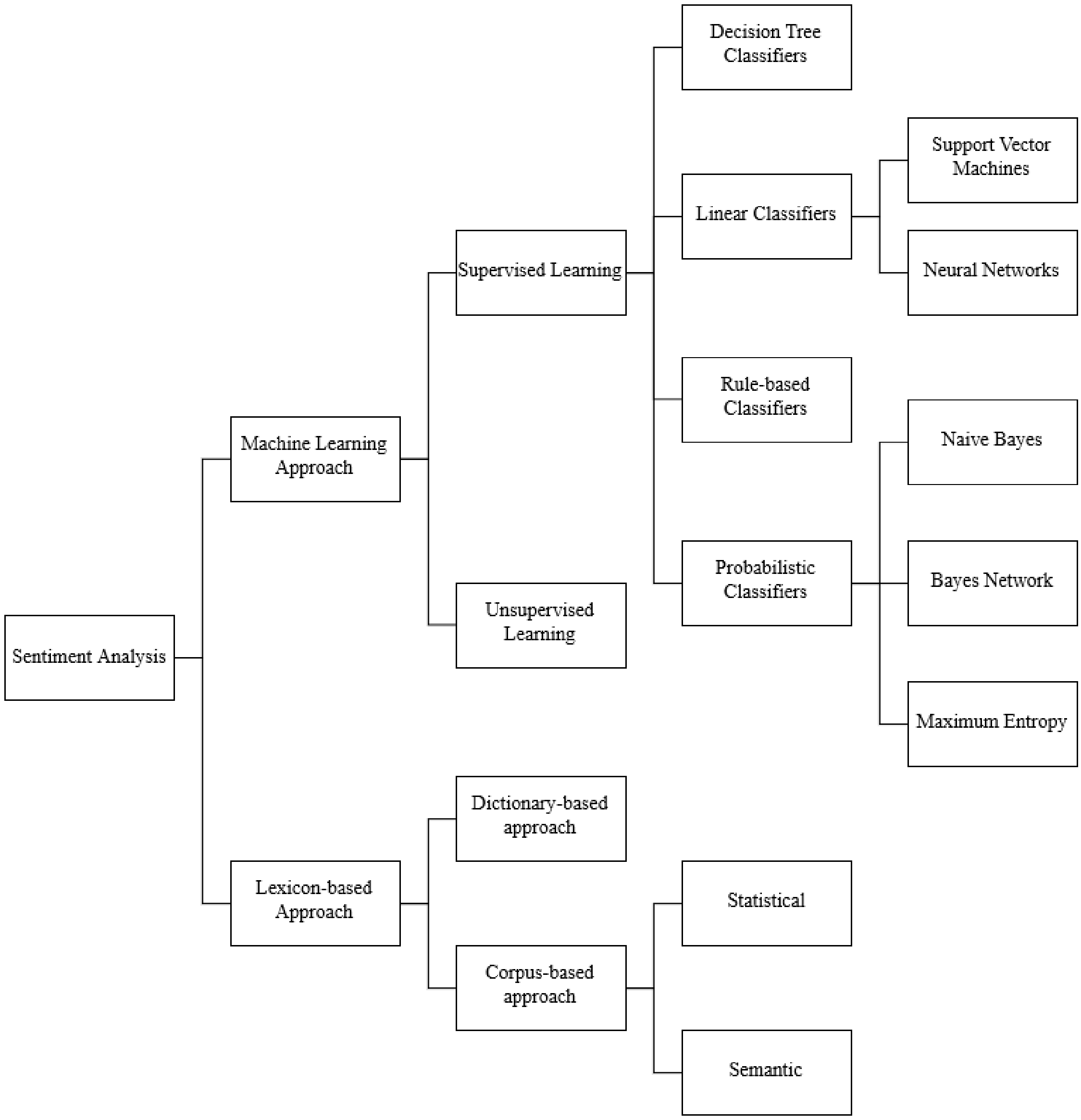

Figure 3.1: Sentiment Classification Techniques Used in SA [MHK14]

He proposed a system and to give a better explanation and draw a diagram that is illustrated in Figure.4. In the following subsections, we show the main steps of the each component respectively, including Data gathering, Data pruning, Mapping, Sentiment Classification, and User Behavioural Analysis.

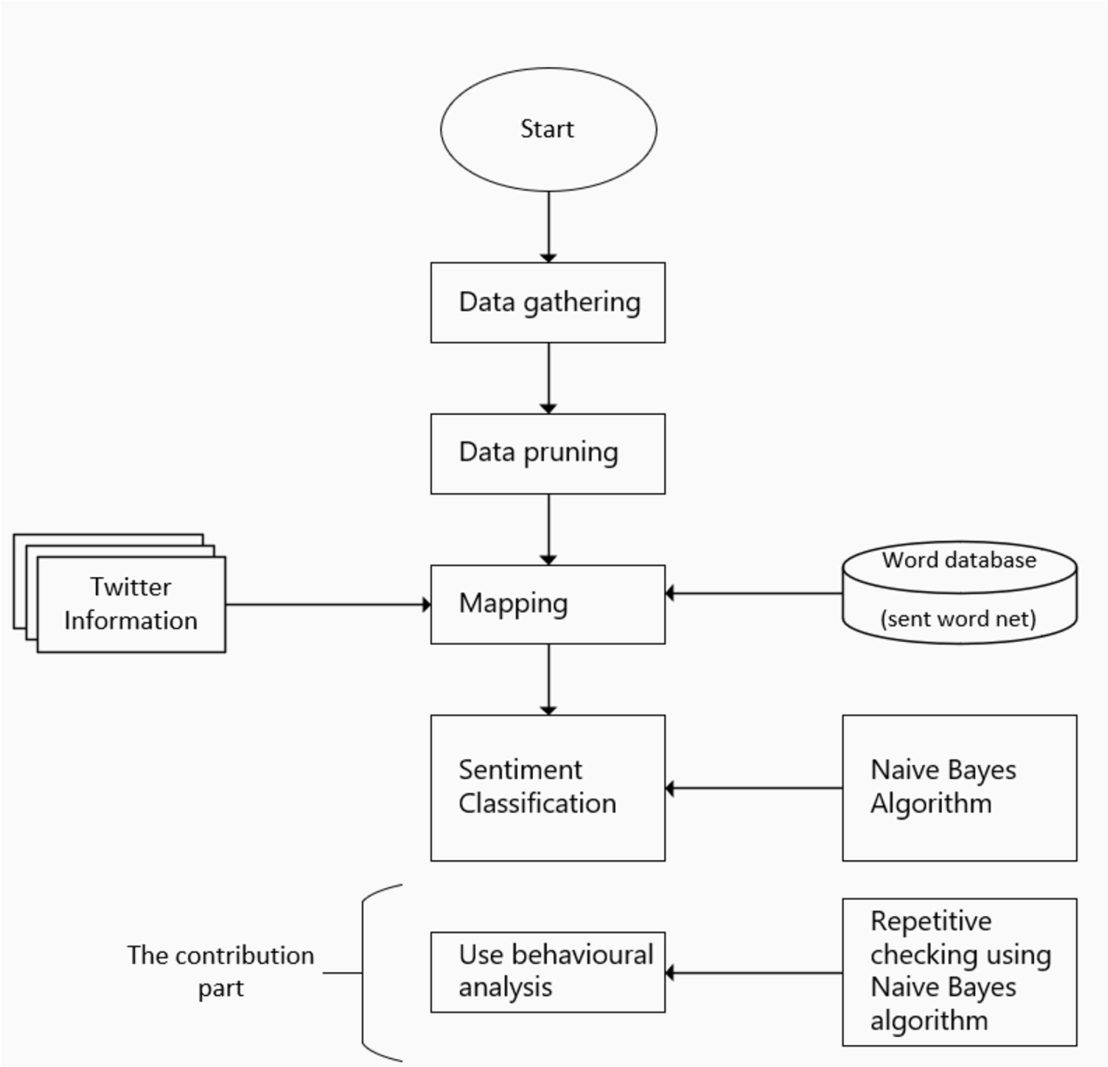

Figure 3.2: Proposed system development diagram [Isk17]

- **Data gathering:** The used data source was Twitter as it is the main social network for terrorism communication, where have been recorded higher issues related to acts of terrorism, even compared to Facebook. So the data was obtained from Twitter streaming API. Than, they used keywords in order to obtain data related to terrorism. And the data is obtained as JSON that is a JavaScript Object Notation [Isk17].

- **Data pruning:** After the data is obtained it needs to be preprocessed in order to normalize it. In this step, they have removed URLs, @tags, hashtags, uppercase and lowercase, spelling errors and etc. [Isk17].

- **Mapping:** They chose to use SentiWordNet [ES06], that is a collection of thousands of English words which have been attributed to a positive or negative score. Using this the sentences obtained will be compared and calculated the score. Since the word alone is not enough to make a decision, the total score will be calculated based on the context of the sentence [Isk17].

- **Sentiment Classification:** In this step sentences were categorized into the classes of positive, negative or neutral. They have used Naive Bayes because it is mostly used on SA. Bayes theorem was used to predict the probability that a given feature set belong to particular label. The Naive Bayes will classify the statement to positive, negative or neutral based on the sentiment score result [Isk17].

- **User Behavioural Analysis:** In this step, they used Snapbird tool [sna] to track the previous tweets of a certain user. When tweets became classified to their polarity based on the sentiment score, all three classes were checked repetitively and then on the second checking tweets on each category were compared with their tweets history.The purpose of the repetitive checking is to get better results on the understanding if tweets are leading towards terrorism or not [Isk17].

In case of the score being negative, after the repetitive checking and resulted on the same class it could be concluded that the account holder might be leading towards acts of terrorism.
The purpose to check users previous tweets is to analyse their tweets patterns. As mentioned above, the user can be an humorist or just making a joke about some terrorism so the pattern of his tweets can be related to jokes. The same can be done on the other side, in case of the user being serious and in risk of radicalisation.
They proposed to use Naive Bayes approach after a comparative analysis. It is clear that Naive Bayes has been proven and had a potential to be implemented. In this paper, Bayes theorem is being applied to predict a class for any giver text from tweets.
The formula 3.1 used by Toit, J.D. [Toi15]:

$$P(label|features) = \frac{P(label)P(features|label)}{P(features)} \quad (3.1)$$

Analysing the formula we get that *P(label)* is the class(positive/negative/neutral) of the tweets while *P(features)* are the tweets themselves. So *P(label|features)* is the result of the application of the technique. Therefore, by this we get the formula 3.2:

$$P(positive|tweet) = \frac{P(positive)P(tweet|positive)}{/P(tweet)} \quad (3.2)$$

The procedure has to be repeated to all three categories. At the end, the highest ranked class is chosen to label the document.
The main conclusion to this paper is that all the elements that were added like user behavioural represent an increase to the accuracy of the SA.

3.2.3 Learning with Hashtags

In this subsection we will analyse a paper named "Learning Emotion Indicators from Tweets: Hashtags, Hashtag Patterns, and Phrases" [QR14]. Furthermore, after the initial training that is needed, the system will do the rest and if it goes well, once the final results depend on the training, it can be applied to any language, which turns the system into a language independent system.
It is usual for users of Twitter to express an emotional state by using hashtags (e.g., **#inlove**, **#hatemylife**). There are some hashtags that consist of a simple word like **#faith**, some are formed by multiple words like **#faithinhumanityrestored** or it can even be a creative spelling for example **sk8** or **cantwait4tmrw**. There are several reasons for this difficulty, but mainly hashtags are constantly being created through a number of endless open combinations.
Qadir and Riloff research [QR14] takes in consideration three tweet indicators: hashtags, hashtag patterns and phrases. It is linked to one of five emotions: Affection, Anger/Rage, Fear/Anx-

iety, Joy or Sadness/Disappointment. They chose to create a bootstrapping framework learning emotion hashtags and general hashtag patterns. And they also harvested emotion phrases from the hashtags and hashtag patterns for contextual emotion classification.

One of the methods used by them is finding common prefix in hashtags. For example, **#angryatlife** and **#angryattheworld** have the same prefix **angry at**, so this suggests an emotion of Anger. Consequently, the specific hashtags are generalized into hashtag patterns that will match hashtags with the same prefix.

A key challenge for the area on this approach to the problem is when we found the same prefixes on hashtags with different emotions, because it cannot result always on the same emotion, as it is explained on Figure.5. In the image we have the example that even if the prefix **angry** includes all three of the hashtags not all of them represent an emotion of anger: the **#angrybirds** does not transmit any emotion at all because it only represents a video game.

In the same line of thought we can extract emotion phrases from hashtags. The example given is having the hashtag **#lovelife** that is associated with Joy. We can harvest the phrase **love life** from the hashtag and use it to recognize emotion in the tweet. The problem here is that, unlike hashtag, the word surrounding a phrase must also be considered in order to obtain tweets emotion. As an example, if we have a negation it would toggle the polarity of the tweet (e.g., **don't love life** may suggest Sadness instead of Joy).

3.2.3.1 Learning Hashtags

For these research, they have used and collapsed Parrot emotion taxonomy ([Par01]) into only five emotions that occur more often on tweets and are easily distinguishable from each other, being them:

- Affection;
- Anger/Rage;
- Fear/Anxiety;
- Joy;
- Sadness/Disappointment.

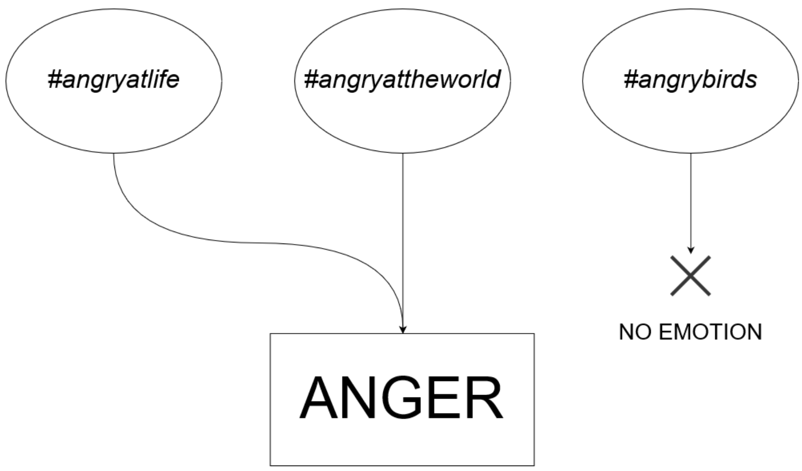

Figure 3.3: Diagram explaining that even with the same prefix it might not transmit the same emotion

And for the cases where there was not found any emotion at all or if it would not feet in any of the chosen there was created a class "None of the above". For every of the five classes they identified five common hashtags strongly associated with the emotion and used them as "seeds".

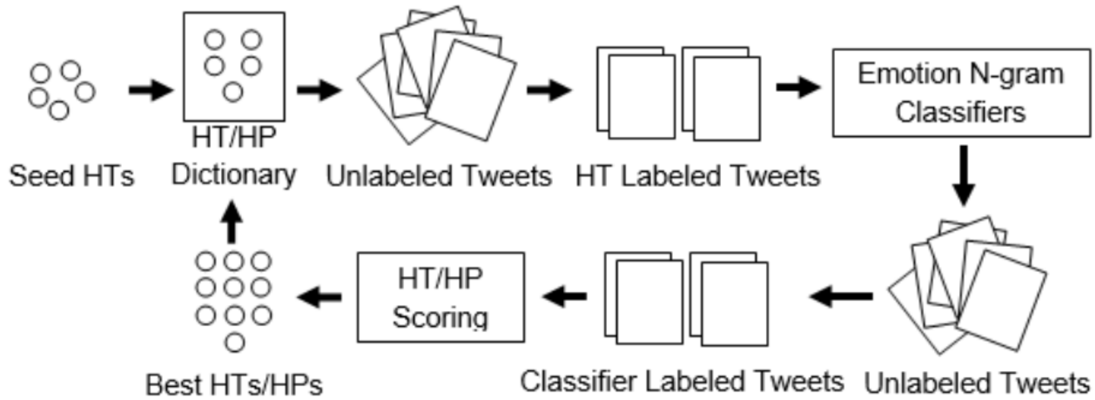

Figure 3.4: Bootstrapped Learning Diagram used by Qadir and Riloff. (HT = hashtag; HP = hashtag pattern) [QR14]

The bootstrapping frame work is represented on Figure.6. Watching the diagram we observe the process begins by tweets containing the seed hashtags and labelling them with the corresponding emotion. There were collect 323,000 tweets with at least one of the seed hashtags. Then were also collected, with Twitter Streaming API, 2.3 million unlabelled tweets with at least one hashtag (1.29 hashtags per tweet and 3.95 tweets per hashtag, both in average).

The labelled tweet would then be used to train a set of emotions classifiers. For each emotion class one logistic regression classifier were trained.

After that, each emotion classifier was applied to the unlabelled tweets. For each emotion, they collected the tweet classified as that same emotion and extracted the hashtags from those tweets to create a candidate pool of hashtags for that emotion. Then the candidate hashtag is scored by computing the average probability of the same emotion above, obtained from the logistic regression classifier, over all the tweet containing the candidate hashtag. Then, from the unlabelled tweets all tweets were added with one of the learned hashtags to the training instance an the process would continue.

In order to reduce the number of candidates, they have discarded hashtags that appear less than ten times, the ones with only one character and the ones with more than twenty times. For us, this should be done in another way because this is a criteria that should be adapted to the objective. In our case, it is important that the hashtag appears more times as possible in order to find collective radicalism and maybe obtain hashtags with more than two characters and no maximum limit.

3.2.3.2 Learning Hashtag Patterns

For the start, author expand the hashtag into a sequence of words using an N-gram based word segmentation algorithm [Nor08]. There was used a Prefix Tree data structure to represent all possible prefixes of the expanded hashtag.

Then, the tree is traversed and all possible prefix are considered as candidates as hashtag patterns. After that, each pattern is scored the same way it was done with hashtags. The way this emotion indicator is obtained is similar to the above, but at the end all tweets have been added with hashtags that match one of the learned hashtag patterns to the training instance and the process continues.

3.2.3.3 Creating Phrase based Classifiers

The last type of emotion indicators that aims to acquire are emotion phrases. Right at the end of the bootstrapping process, the word segmentation algorithm that was applied above was used to all hashtags and hashtag patterns in order to expand them into phrases.
It is assumed that the obtained phrase has the same emotion as the original hashtag. Nevertheless, it will have low precision, because it was already said that the contexto must be taken into account.
This time a logistic regression classifier was trained for each emotion, which classified a tweet about its emotion based on the presence of learned phrase for the emotion in case, as well as a context window of size six around the obtained phrase, three for each side of the phrase.

Thus, in conclusion, once this is applied to hashtags and that hashtags gain form in every language, it can be applied to any language, so it is language independent, as it does not depend on any specific corpus, it can be applied by anyone to its native language.
Even though this system needs to be supervised, it is weakly supervised so it can be a help in the way to build a unsupervised system.

3.2.4 How to be Unsupervised

An unsupervised approach is an approach where training inputs are not necessary for the system to discover the target point of the learning. Besides that, it would need to do the learning by the system itself and there is no need for any human supervision or intervention on the system to achieve the learning goals. So the only time a human is needed for a system like this is when the client needs to change or add functionalities in the system, and so the developer will add it and after that the system will again be doing the job itself.
[QR14] presented a weakly supervised system and it is hard to use the method as totally unsupervised but is not impossible. We aim to use the information from this paper to build our approach and create an unsupervised system and instead of using hashtags we will make this in a probabilistic approach, which give us room to make the approach unsupervised.
Once we will be using probabilistic methods to analyse the sentiment of the tweet it will be easier to find a way of automatize all the system and allow it to work without human supervision. Besides saving time for the human, it saves the analysis of becoming influenced by the human. When we talk about an unsupervised we must not expect it to do more than the job that it is designed for. For example, if we created a software that it is listening to all tweets of a user and summarizing it without supervision, we must not expect it to make a SA by itself. It is doing the job for what we have created it with no supervision, but without developing the SA part it would create it, by itself. Maybe in a near future a computer can create its own programs, but not for now.

3.2.5 How to be Language-Independent

An approach that is language independent was introduced by Dias et al.[DM], where the criteria for the approach to be language independent is that although it is applied to different languages the results keep being satisfactory and the experiment values represent the reality in a viable way.

Besides that, in order for the approach to be language independent when it is applied to another language, it must not require any changes other than the input data, and if the system needs training it will be needed to give information on the language it will be applied to.
As a second objective we aim that the system would do the work for any language it founds. For example, if we do get tweets from the Streaming API, having as a criteria the geo-localization as Portugal there would be tweets not only on Portuguese but probably on other languages. So we would like our system to analyse the tweet regarding the language it is written.
Another point we aim to achieve is take in count the geo-localization of the tweet we are analysing because by having two equal tweets but from different regions of a country it can modify the meaning of the tweet. As we can conclude from [BDS14], the language of a user might change according to the social context the user is included.
Besides that, we should be taken into account that with the same tweet different users from different regions, and even if the tweet is a little older it might not have the same emotion, sentiment or meaning.

3.2.6 Specific possible methods and tools to be used

As we will use the Python programming language we will use NLTK for these language. "NLTK is a broad-coverage natural language toolkit that provides a simple, extensible, uniform framework for assignments, demonstrations and projects. It is thoroughly documented, easy to learn, and simple to use. NLTK is now widely used in research and teaching" ([BL04]).
We also have in mind the use of another tool existing on this toolkit we will try to develop a method based on Naive Bayes but one that is unsupervised, and if we use it we will be closer to the to main focus on our work: language independence and unsupervised approach. We have other option in order to develop our project.
According to [SD18], the following points explains possible machine learning techniques that we might use for our project :

- Naïve Bayesian: "The concept of this technique classifier is based on Bayes theorem with independence assumptions between predictors, as we saw in of our related works. A Naïve Bayesian model is easy to build, with no complicated iterative parameter estimation which makes it particularly useful for very large datasets. Despite its simplicity, the Naive Bayesian classifier often does surprisingly well and is widely used because it often outperforms more sophisticated classification methods" ([SD18]) the formula used for this technique is seen on equation 3.3 ;

$$P(A \mid B) = \frac{P(B \mid A)\, P(A)}{P(B)} \quad (3.3)$$

- Neural Networks: "Are modelled after the biological neurons in brain structures. The individual neuron models may be combined into various networks made up of many individual nodes, each with its own set of variables. These networks have an input layer, an output layer, and one or more hidden layers. The hidden layers provide connectivity between the inputs and outputs. The network may also have feedback, which will take result variables and use them as input to prior processing nodes" [SD18]:
  - Modelled after the human brain
  - Experimentation and marketing predated theory

- Considered the forefront of the AI spring
- Suffered from the AI winter
- Theory today still not fully developed and understood
- A network of interconnected functional elements each with several inputs/one output

$$y(x_1, \ldots, x_n) = f(w_1 x_1 + w_2 x_2 + \ldots + w_n x_n) \quad (3.4)$$

- w_i are parameters of equation 3.4;
- f is the activation function of equation 3.4;
- Crucial for learning that *addition* is used for integrating the inputs

- K-Nearest Neighbour: "It does not have any learning phase because it uses the training set every time a classification performed Nearest Neighbors search(NN) also known as proximity search, similarity search or closest point search is an optimization problem for finding closest points in metric spaces. K nearest neighbour is applied for simulating daily precipitation and other weather variables" [SD18];
- Decision Trees: "The decision tree is one of the popular classification algorithms in current use in Machine Learning. Decision tree is a new field of machine learning which is involving the algorithmic acquisition of structured knowledge in forms such as concepts, decision trees and discrimination nets or production rules" [SD18].

Each type of approach has its own advantages and disadvantages.
For Naïve Bayesian technique it is fast to train and to classify but on the other hand it assumes independence of features which can make it unusable for an approach [SD18].
In the case of Neural Networks they are not sensitive to irrelevant features contrary to the last technique. This is also realized as specialized hardware systems. And it is really useful for network learning. On the other side, this is too much of a black box technique and it is not probabilistic, which is not what we aim with our project [SD18].
K-Nearest Neighbour is a good approach once we get data from Twitter and it can be quite noisy data. This technique is robust to noisy training data and even if there is a large amount of training data this technique will still be effective. On the other hand, K value needs to be determined which is not easy in our problem. Furthermore, is has a high computation cost [SD18].
Finally, the Tree Decision technique, which offer an easy way to understand and interpret calculations and can always be used with other decision techniques. But, depending on our objectives, it can get very messy and complex and we must not forget that the gained information is biased [SD18].
All the techniques above are supervised. This was taken into account because it can be a start point and maybe we can complete it in order to turn it into unsupervised techniques.
As native unsupervised techniques we have Lexicon based methods, dictionary based methods and Corpus based methods.
Lexicon based approach uses insights obtained on the ground of polarity of words composing a sentence. With this method, we can create a categorical evaluation (Positive, Neutral, Negative) or we can calculate a score [JSD16].
In dictionary based techniques the idea is to first collect a small set of opinion words manually with known orientations, and then to grow this set by searching in the WordNet dictionary for

their synonyms and antonyms. The newly found words are added to the seed list. The next iteration starts. The iterative process stops when no more new words are found [VT13]. Corpus based techniques rely on syntactic patterns in large corpora. Corpus-based methods can produce opinion words with relatively high accuracy. Most of these corpus based methods need very large labelled training data. This approach has a major advantage that the dictionary-based approach does not have. It can help find domain specific opinion words and their orientations [Nas08].

3.3 Word Embeddings

As we want to use word embeddings on our work in order to extend ExtremeSentiLex into a second version with more words were we can find the word in different verbal forms and also find related words we will use word embedding by similarity. Being so, we want to learn about this area so we will explore and quote some related works on this area.

Representation learning is a long-standing problem in NLP. The main motivation is to abstract away from the surface forms of a text piece, e.g., words, sentences and documents, in order to alleviate sparsity and learn meaningful similarities, e.g., semantic or syntactic similarities, between two different pieces of text. Learning representations for words as the atomic components of a language, also known as word embedding.

Word Embeddings are distributed word representations for a set of language modelling [MMS99] and feature learning [DHS12] techniques from NLP. Two of the best known model to generate word embeddings are Word2Vec [MCCD13] , which we used in our work and Global Vectors (GloVe) [PSM14]. This two methods learn low-dimensional word representation for words by taking into account word pair co-occurrences, captured in a word context matrix. [LQY+18] Both methods have shown good results. To choose one of them we need to have in mind that Word2Vec takes more time to train once it computes the matrix line by line and GolVe takes more memory to get the work done.[Got16]

One of the most popular model for word embeddings is neural network-based language models. The Word2Vec model proposed by Mikolov et al. is an embedding model that learns word vectors via network with a single hidden layer. We find some successful works in learning semantic word representations is employing global matrix factorization over word-word matrices. These models have been demonstrated to be effective in short text similarity[KDR15] and document model estimation.

We will speak about Word2Vec once the file we will be created using Word2Vec applying many news from Google News. The best thing about Word2Vec is it is simple and accessible, the code is open and anyone can download it.

Word2Vec is an unsupervised system for determining the semantic distance between words. Since the publication of the paper in 2013 it has been cited already 1328 times. This reveals the usability of this tool. When using Word2Vec it is assigned a number between 0 and 1, indicating the semantic distance between two words.

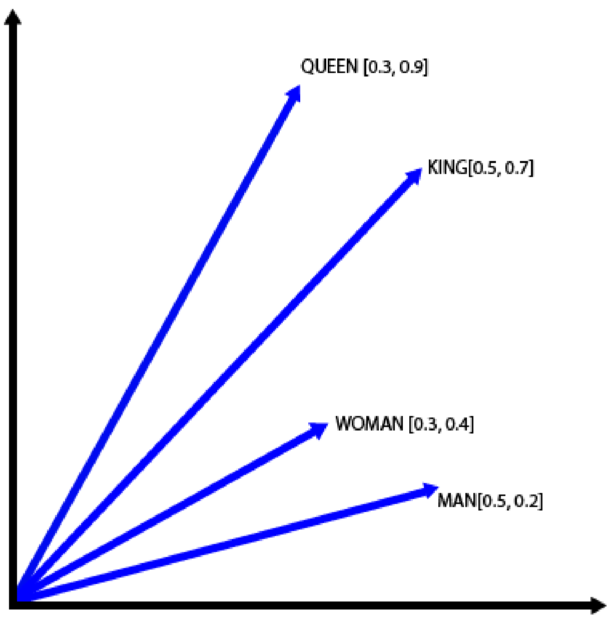

Figure 3.5: Example of words vectors

In the image 3.5 we have 4 vectors where we can understand how vectors distance works. So has we can calculate if we have queen = $\vec{u}$=<0.3, 0.9> and king = $\vec{v}$=<0.5, 0.7> as showed in equation 3.5:

$$\frac{\vec{u} \cdot \vec{v}}{\|\vec{u} * \vec{v}\|} \equiv \frac{0.3 * 0.5 + 0.9 * 0.7}{\sqrt{0.3^2 + 0.9^2} * \sqrt{0.5^2 + 0.7^2}} \equiv \frac{0.15 + 0.63}{0.95 * 0.86} = 0.96 \quad (3.5)$$

For queen and woman can be seen in equation 3.6:

$$\frac{\vec{u} \cdot \vec{v}}{\|\vec{u} * \vec{v}\|} \equiv \frac{0.3 * 0.3 + 0.9 * 0.4}{\sqrt{0.3^2 + 0.9^2} * \sqrt{0.3^2 + 0.4^2}} \equiv \frac{0.09 + 0.36}{0.95 * 0.5} = 0.95 \quad (3.6)$$

For queen and man can be seen in equation 3.7:

$$\frac{\vec{u} \cdot \vec{v}}{\|\vec{u} * \vec{v}\|} \equiv \frac{0.3 * 0.5 + 0.9 * 0.2}{\sqrt{0.3^2 + 0.9^2} * \sqrt{0.5^2 + 0.2^2}} \equiv \frac{0.15 + 0.18}{0.95 * 0.54} = 0.64 \quad (3.7)$$

The distances are different from queen to the other words, but it is seen that the man is less similar. From this what can be done is chose the two most similar words in order to expand the vocabulary, in this case, **queen** would add to the vocabulary **king** and **woman**, and would not add **man**. It can also be added the two words just with a distance similarity higher than 0.9 and would also add **king** and **woman**, if it was lower than 0.9 **man** would be added.

3.4 Conclusion

In this chapter, we found much information about what has been done in the areas to come close to a solution. Despite the fact, there is not much work that we could use to join all the areas we are trying to join, we were able to collect information to create the state-of-the-art for this specific area.

The area of extremism and/or radicalisation online doesn't have much previous work. Although, there is plenty information on SA on computer science area. There is also information on roots

of radicalisation and extremism. Being so, we will work joining the areas. Once we know how to analyse the sentiment of online posts, we only need to apply the knowledge on radicalisation and extremism we got from this section.
After we analyse the posts, it would be necessary to get the information of that users who post many extreme posts and give them some attention. Get a record from every post, maybe cross multiple platforms having the same user.

Chapter 4

Extremism Understanding

Online social networks (OSNs), such as Facebook, Twitter and Tumblr [1], have become a de-facto platform for hundreds of millions of Internet users. This social platforms and many others, are new kinds of blogging that make it easier to communicate with people all over the world.

At one end the posts or tweet messages enable people to understand constructive views, ideas or share articles, videos or links, etc.

In this study, we present an innovative approach to detect extreme sentiments on social media. We aim to analyse corpora in order to identify extremist posts and used ExtremeSentiLex having a negative and positive polarity. We have designed and developed a prototype application called *Extreme Sentiment Generator* by applying on two lexical resources, SENTIWORDNET 3.0 [BES10] and SenticNet 5 [CPHK18], in order to obtain an extreme lexical resource that contains positive and negative terms, which represents extreme sentiments using statistical methods as described in sections below.

Furthermore, we also analyse the compilation of datasets in order to ensure the accuracy of the approach to find extreme sentiments on social media posts. We classified five corpus already used and studied by the scientific community, all of them constituted by posts found on social media and some with only social network posts. We use a confusion matrix to calculate recall, precision and accuracy of our application performance.

This chapter is organized as follows: Section 2 briefly introduces the background; Section 3 discusses the related work; Section 4 demonstrates our methodological approach to collect the lexicon of extreme sentiments; Section 5 shows experimental setup; Section 6 presents and discusses the results; Section 7 concludes the chapter.

4.1 ExtremeSentiLex

In this section we present methodological approach to collect a lexicon of extreme sentiments through two datasets: **SENTIWORDNET 3.0** and **SenticNet.5**. Figure 4.1 show the overall process of extreme sentiment collection.

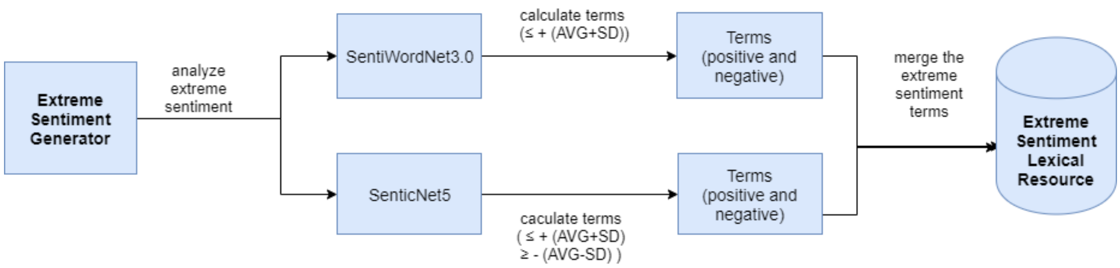

Figure 4.1: Extreme sentiment collection process.

[1] https://www.tumblr.com/

4.1.1 Used Lexical Resources

SENTIWORDNET 3.0 is developed using the automatic annotation of all WORLDNET synsets, using the notions of 'positivity', 'negativity' and 'neutrality'. Each synset has three numerical scores, which indicate the terms as positive, negative, and objective (i.e., neutral): e.g., *majestic score: 0.75 (positive term), invalid 0.75 (negative)*. In [BES10], SENTIWORDNET 3.0 is used as a base for the development of extremism lexical resource, an enhanced lexical resource that is developed to be used as support for sentiment classification and opinion mining applications [PL08]. This lexical resource is publicly available for research purposes.
SenticNet 5 [CPHK18] encodes the denotative and connotative information commonly associated with real-world objects, actions, events and people. It steps away from blindly using keywords and word co-occurrence counts and instead relies on the implicit meaning associated with common sense concepts. Superior to purely syntactic techniques, SenticNet 5 can detect subtly expressed sentiments by enabling the analysis of multi-word expressions that do not explicitly convey emotion but instead are related to concepts that do so. An example of SenticNet 5 datasets is: *favourite 0.87 (positive), worry -0.93 (negative)*.

4.1.2 Defining Extreme Polarity

The first phase to collect extreme sentiments is to define the extreme polarity for the terms. The goal of this phase is to establish a metric to classify the terms that have extreme scores both for positive and negative. Referring to Figure 4.1, we developed a python application for *Extreme Sentiment Generator (ESG)* that performs a number of operations i.e., calculates the average and standard deviation of terms from the original lexical resources, filters and saves them in the new lexical resource. We define two general equations (Equation 4.1 and Equation 4.2) in **ESG** to categorize both positive and negative terms respectively. Since each dataset has a different classification of terms, we use either one equation or both equations to identify extreme positive and negative sentiments, where T_p refers to positive terms, and T_n to negative terms.

$$EP = Average + StandardDeviation < T_p \quad (4.1)$$

$$EN = Average - StandardDeviation > T_n \quad (4.2)$$

Afterwards, we process both datasets, one by one as follows.

SENTIWORDNET 3.0: This dataset has three categories for terms: 'positive', 'negative' and 'neutral'. In this dataset, the score for both positive and negative terms are in a range of $[0, 1]$. First we filter this lexical resource and obtain only positive and negative terms separately. Next we use only Equation 4.1 for identifying extreme negative and positive terms.
With the calculation using **ESG**, we obtained the following output:

```
Average for positive: 0.366, Standard Deviation for positive: 0.211
Extreme polarity for positive terms: 0.577
Average for negative: 0.412, Standard Deviation for negative: 0.230
Extreme polarity for negative terms: 0.642
```

The output shows that positive extreme polarity is 0.577, while negative extreme polarity is 0.642. Consider the sample output example terms of SENTIWORDNET 3.0 generated by **ESG**:

```
ultrasonic 0.375 (non positive extreme), selfless 0.875 (positive extreme)
thrash     0.125 (non negative extreme), abduction 1 (negative extreme)
```

Therefore, *selfless* is a positive extreme as $0.577 < 0.875$, while *ultrasonic* is not. Similarly, *abduction* is negative extreme $0.577 < 1$ and *thrash* is not. Moreover, we discarded all non-positive and non-negative extreme terms from our obtained lexical and save this result in a file.

SenticNet 5: For dataset, it is easier to find the extremes as each term has one score in a range between $[-1, 1]$ i.e., negative and positive. To calculate the extreme polarity for both terms, we rely on both Equations 4.1 and 4.2.
The output of the result is as follows:

```
Average for positive: 0.504, Standard Deviation for positive: 0.362
Extreme polarity for positive terms: 0.866
Average for negative: 0.616, Standard Deviation for negative: 0.306
Extreme polarity for positive terms: -0.922
```

Again, only positive terms with intensity greater than 0.866 are considered as positive extremes, and negative terms with intensity lower than -0.922 are taken as negative extremes. Consider the following sample example output of extreme positive and negative terms:

```
grace 0.79 (positive non extreme), pioneer 0.97 (positive extreme)
anemic -0.91 (negative non extreme), traffic -0.97 (negative extreme)
```

Again, we discarded all non-positive and non-negative extreme terms from our obtained lexical and save this result in a file.

4.1.3 Generating ExtremeSentiLex

In this phase we generated our final ExtremeSentiLex. To achieve it, we merge both files obtained from SENTIWORDNET 3.0 and SenticNet 5. In SENTIWORDNET positive and negative extremes lay in $[0, 1]$ interval, while in SenticNet scores range from -1 to 1, for negative (< 0) and positive (> 0) extremes. To uniform the scales, we multiply all the negative terms of SENTIWORDNET 3.0 by -1 to obtain a range in $[-1, 1]$. Then, we merge both files, remove all duplicate terms by considering the ones with highest score and create the final file of *Extreme-SentiLex* as shown in Figure 4.1. The final result is a text file with two columns: the term and its corresponding intensity of the term. Below is a sample output of terms with score:

Term	Score	Term	Score
absolutely	+0.88	accept	+0.93
acknowledgeable	+0.95	acne	-0.96
actively	+0.95	adroitness	+0.88
agent	+0.91	agoraphobic	-0.95
alright	+0.88	amuse	+0.92

After we create our first version of ExtremeSentiLex, we decided to use word embeddings in order to extend our lexicon. We used a file with word embeddings extracted from Google News and we calculated the ten closest terms to every term of ExtremeSentiLex and only use the term if semantic distance was not lower than 0.5. After that we needed to apply filters on the terms we got from there. The filter we decided to apply was if the term was founded in the original lexical resources (SenticNet5 or SenWordNet3.0), then we excluded that term. Being so, the first version of ExtremeSentiLex has 398 terms and the second version has grown to 1121 terms. It is a growth of, approximately, 181% of the original ExtremeSentiLex. This work was made using a program written in python.
Our hypothesis is by using the extended version of ExtremeSentiLex our result on chapter 5 on the first section will be improved. Once we have more terms we will find more terms where the context was equal but were not recognized as extremes, sometimes the only thing different is the verbal form, and on the extended version were included different verbal forms of the same word, but we also have terms which are synonyms.

4.2 Detect Extreme Sentiments

4.2.1 Datasets

The use of different datasets by researchers for SA is an important aspect to analyse the users sentiments. Here we present some of the important datasets that are being used for SA in different studies. In here, five datasets consist of posts and/or tweets messages, and two datasets that are constituted by lexicon related to sentiments.
Sentiment 140 [GBH09] is a dataset, which consists of two CVS files, one for test and another for training. Sentiment 140 provides one sentiment value per tweet on a scale from 0 (negative) to 4 (positive). For better comparison, values are converted to obtain three sentiment categories: positive, negative, and neutral. Sentiment 140 is used in the work [FBSH15], which presents tweets SA using Sentiment 140 and SentiStrength [2] on a large representative set of research papers that specifically adapt different techniques to education articles distributed on Twitter.
Rt-polaritydata [PL05] is another dataset that contains two files: one for positive text and the other for negative reviews. There are 2000 comments, processed and classified in two different categories. Comments usually consist of a number of sentences, but the users opinion will be identified at the sentence level, which later identified overall comments opinion. The found collection consists of two files, one for each set of 5331 positive opinions and negative opinions, containing one sentence per line, making it easy to process [SB+12].
Twitter for Sentiment Analysis (T4SA) images dataset [VCC+17] consists of both the textual and multimedia data. In [VCC+17] authors have gathered data from Twitter by means of a streaming crawler. According to authors the process of data collection took like six months i.e., from July to December 2016 and then used for visual SA evaluation. The study concludes that

[2] http://sentistrength.wlv.ac.uk/

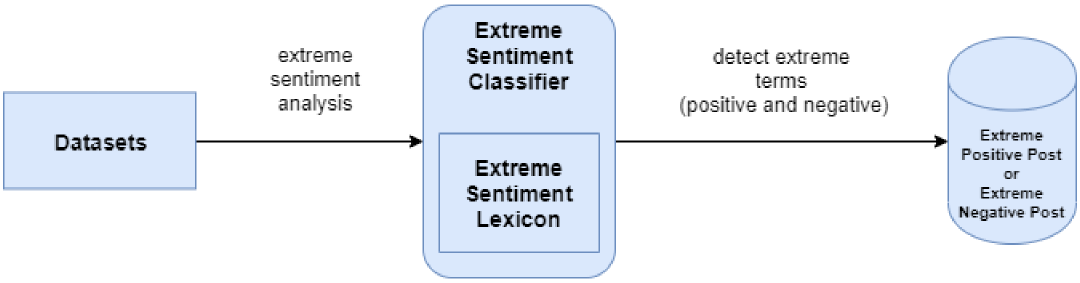

Figure 4.2: Process of testing of Extreme Sentiment Classifier

the approach is effective for learning visual sentiment classifier. T4SA dataset and the trained models are publicly released for future research and applications.
TurntoIslam [AIL13b] and *Ansar1* [AIL13a] are datasets having posts that are organized into threads, which generally indicate topic under discussion. Each post includes detailed metadata e.g., date, member name, etc. As announced on the forum, this is an English language forum having as goal 'Correction of common misconceptions about Islam'. Radical participants may occasionally display their support for fundamentalist militant groups as well.
With all existing approaches and systems, the focus of our work is to find extreme sentiments representation on social media and social networks. Our work is based on using existing datasets (dictionaries) that contains an intensity score for terms. To achieve it, we used two datasets, SENTIWORDNET 3.0 and SenticNet5, then we defined a measure to find only the most intense terms that reflect extreme sentiments on a set of numerous terms with our developed approach presented in section 4.1.

4.2.2 Experimental Setup

To evaluate the performance of our application *Extrem Sentiment Classifier* in terms of accuracy, we choose three social media corpus: TurnToIslam [AIL13b], Ansar1 [AIL13a], rt-polaritydata [PL05] and two social network corpus: T4SA Images Dataset [VCC$^+$17] and Sentiment 140 [GBH09] for experimentation. We developed another application having *Extreme Sentiment Classifier* (**ESC**) having ExtremeSentiLex embedded in it. The real purpose of this experimentation is to analyse whether **ESC** is able to classify the extreme positive and negative terms from the five different databases or not. We further use confusion matrix [Tin17] to describe the performance of our classification model by computing recall and precision for extreme positive, negative terms, and overall accuracy from the results obtained from experimenting corpus.
Figure 4.2 depicts overall process of applying **ESC** on the five datasets. **ESC** is applied on datasets to find terms inside posts, the score of each term found is then added. Later we have defined Equation 4.3 to identify the posts by showing extreme sentiments (positive and negative). We consider that only such posts as extreme, i.e. those that satisfy the equation. In the equation bellow, PS refers to the sum of all positive terms score and NS refers to the sum of all negative terms score.

$$EXTREME: \; |PS - |NS|| > \frac{PS + |NS|}{2} \quad (4.3)$$

Next we define three conditions for all posts and declare the result accordingly.

```
IF  [PS > |NS|] && [EXTREME]
STATE `The post is classified as Extreme Positive'
ELSE IF [PS < |NS|] && [EXTREME]
STATE `The post is classified as Extreme Negative'
ELSE
STATE `The post is classified as Inconclusive'
```

Consider the following example for extreme positive from the datasets:

> *Example: Since when does #alcohol equal #**happiness**? I know many people that started drinking; haven't been **happy** since.*

Where the terms and their scores are:

stupid happiness +1.0, happy +0.89

Above we see a tweet with two words that represents extreme positive sentiment so we sum the scores and apply the algorithm:

$$|PS - |NS|| > \frac{PS+|NS|}{2} \Leftrightarrow |1.89 - 0| > \frac{1.89+0}{2} \Leftrightarrow 1.89 > 0.945$$

Being true, it means this is a positive extreme sentiment tweet. Consider an example for negative extreme:

*Example: Yesterday, I had to pay for all the **stupid** things I learned to say.*

Where the terms and their scores are:

stupid -0.93

Above we see a tweet with one word that represents extreme negative sentiment so we sum the scores and apply the algorithm:

$$|PS - |NS|| > \frac{PS+|NS|}{2} \Leftrightarrow |0 - 0.93| > \frac{0+0.93}{2} \Leftrightarrow 0.93 > 0.465$$

Being true, it means this is a negative extreme sentiment tweet. An example of inconclusive term is shown below:

Example: Hustlers don't sleep, we nap!

$$|PS - |NS|| > \frac{PS+|NS|}{2} \Leftrightarrow |0 - 0| > \frac{0+0}{2} \Leftrightarrow 0 > 0$$

As it is false, the post is recognized as inconclusive.

The choice for this classification method was a result of the study of previous works done in this area. As we want the system to be language independent we choose to work with statistical methods for every step of the work. By doing it, the system is able to work with any language. We created ExtremeSentiLex in English using statistical methods and then we classify the post by using also statistical methods on our ESC.
We also aim to provide ESC for the scientific community along with the other files already mentioned in the first chapter. With this we hope for the toll to be applied all around the globe in order to prevent major problems of extremism and collective radicalisation. For this purpose we will make it available in opensource, so this way everyone can make its changes in order to improve our system.

4.3 Conclusion

On this chapter we showed how ExtremeSentiLex and its extended version were generated and its major strengths and also weaknesses. Furthermore, we also showed how we apply ExtremeSentiLex to detect the extreme sentiments on text.

It was a really long process because, as we were generating a lexicon of extreme sentiment terms, it was necessary to make a human verification in every iteration of the process in order to get the best results possible. We executed the ESG many times always applying improvements on the code. This process took also much time because we aim for the final result to be totally functional and easily used by the community. Once ExtremeSentiLex is unique and innovative tool it can also be improved but it is already functional.

Also, it was necessary to evaluate the original lexicons to find all the needed measures in order to find the terms for the generation of ExtremeSentiLex. This part took also much time once we tried to analyse works done using SenticNet and SentiWordNet, discovering the different ways we could work with this datasets. By doing this effort we were able to build our ESG, but the most important part of the creation of ESG was to chose the best metrics to get the words with the highest sentiment intensity.

For the ESC we needed to choose a statistical based formula to detect extreme posts, after some theoretical hypothesis being test at the begging of the work we choose to use the formula on equation 4.3 once it was the most sustainable one.

Chapter 5

Tests and Validation

In this chapter are showed the results of all our experiments made using ExtremeSentiLex. We made experiments on various datasets with different contents, but mostly with information from social networks and social media. For that, we analysed its polarity on the extremes the datasets rt-polarity, T4SA, Sentiment140, Ansar1, and TurnToIslam.

After showing the results obtained on tables and graphics we make an analysis on results to explain our point of view. We will generate more graphics to compare our approach with and without word embeddings and explain it making an analysis. A comparative discussion on this section is important once we only can know we have good results when we have a comparison term.

We are using two types of datasets: we have some where we can compare sentiment polarity with the extreme polarity we detected and we have others which are more likely to have less inconclusive values once they are datasets obtained from radicalisation forums.

5.1 Tests and Results

In this section, we report the results in the tables 5.1, 5.2, 5.3, 5.4, 5.5, and 5.6 without using word embeddings and tables 5.7, 5.8, 5.9, 5.10,5.11, and 5.12 with word embeddings respectively. The arrangements for each table are different based on the datasets original classification. For example, for **Ansar1** and **TurnToIslam**, we present the count for every classification due to the default arrangement of original datasets i.e., extreme positive terms, extreme negative terms, inconclusive posts, etc. We count such posts that contain extreme words from our lexicon that are not classified as extremes rather them these groups are divided by its polarity i.e., P>N as positives but not extreme and N>P as negative but not extreme.

For the other tables, from the second column, it represents the original classification and defines our extreme sentiments from the second row. In Table 5.6, we calculate all the aggregated measures from confusion matrix [Tin17] i.e., recall and precision for both positive and negative values. It is due to both can be the true positives/negatives or the false positives/negatives. Moreover, we also calculate the accuracy of an entire matrix. In the table where is EP refers to extreme positive posts, EN refers to extreme negative posts, INC refers to inconclusive posts, P refers to positive posts, and N refers to negative posts. The $Total$ value present on the tables 5.1, 5.2, 5.3, 5.4, 5.5, 5.6, 5.7, 5.8, 5.9, 5.10,5.11, and 5.12, refers to the value of posts present on the original dataset.

After analyse the first results obtained a new hypothesis was created: if both lexicon and dataset are balanced the result would be more accurate. To do this, it was needed to delete positive terms to get down to the same number as negative. The deleted terms were the terms with the lowest score. And for the datasets it was chosen B-T4SA which is a subset of T4SA that is balances. The results of this experiment can be find on tables 5.13, 5.14, 5.13 and the result of the application of confusion matrix is on table 5.16 and figure 5.9.

5.1.1 Results for ExtremeSentilex

In this subsection we find the results obtain by the execution ESC by using the ExtremeSentiLex as an input, the other input is the dataset. Bellow we have tables with this results for a better understanding and an easiest reading.

Table 5.1: rt-polarity results

	P	N	Total
EP	971	675	1646
EN	99	183	282
INC	4261	4473	8734
Total	5331	5331	10662

Table 5.2: T4SA results

	P	N	Neutral	Total
EP	82707	10206	37422	130335
EN	1336	8206	1110	10652
INC	287298	160638	591034	1038970
Total	371341	179050	629566	1179957

In the table 5.1 we have the results obtained for the rt-polarity dataset, looking at the numbers the results look good, once we have high number of inconclusive posts as also we will find in the other datasets. As expected once we have a higher number of positive extreme terms than negative extreme terms, the number of negative posts found is lower then the positives ,it is expected the results to be good. In table 5.2 we have a big huge difference between the amount of *EP* and *EN* this is due to the difference between positive an negative posts in the dataset and also to ExtremeSentiLex being unbalanced.

Table 5.3: Sentiment140 training results

	P	N	Total
EP	83827	57912	141739
EN	5258	14986	21244
INC	710915	726101	1437016
Total	800000	799999	1599999

Table 5.4: Sentiment140 test results

	P	N	Neutral	Total
EP	20	11	2	33
EN	1	11	0	12
INC	160	155	137	452
Total	181	177	139	497

As we found on table 5.1 in table 5.3 as it is also balanced we found result very similar but in different amount so the result are expected to be also similar. In table 5.4 we expected the results to be close to the results to 5.3, once it is the test dataset for the training dataset of Sentiment140. And for quick look it gives us a little different result.

Table 5.5: TurnToIslam and Ansar1 results

Datasets	EP	EN	P>N	N>P	INC	Total
TurnToIslam	94110	3583	3842	2503	231300	335328
Ansar1	5882	796	520	491	21803	29492

The analysis of table 5.5 is different from the others, in this table we were trying to find more cases of extremism posts, but due to the large size of the posts our ESC find many terms negative and positive in the same post and in majority of the cases it happens that calculating the number of positive and negative terms on the post the result when we apply our formula might be zero or not extreme.

Table 5.6: Indicators of algorithm efficiency

	Datasets			
	rt-polarity	Sentiment 140		T4SA
		Training	Test	
Recall_{EP}	91%	94%	95%	98%
Recall_{EN}	21%	22%	50%	45%
Precision_{EP}	59%	59%	64%	89%
Precision_{EN}	65%	75%	92%	86%
Accuracy	60%	61%	72%	89%

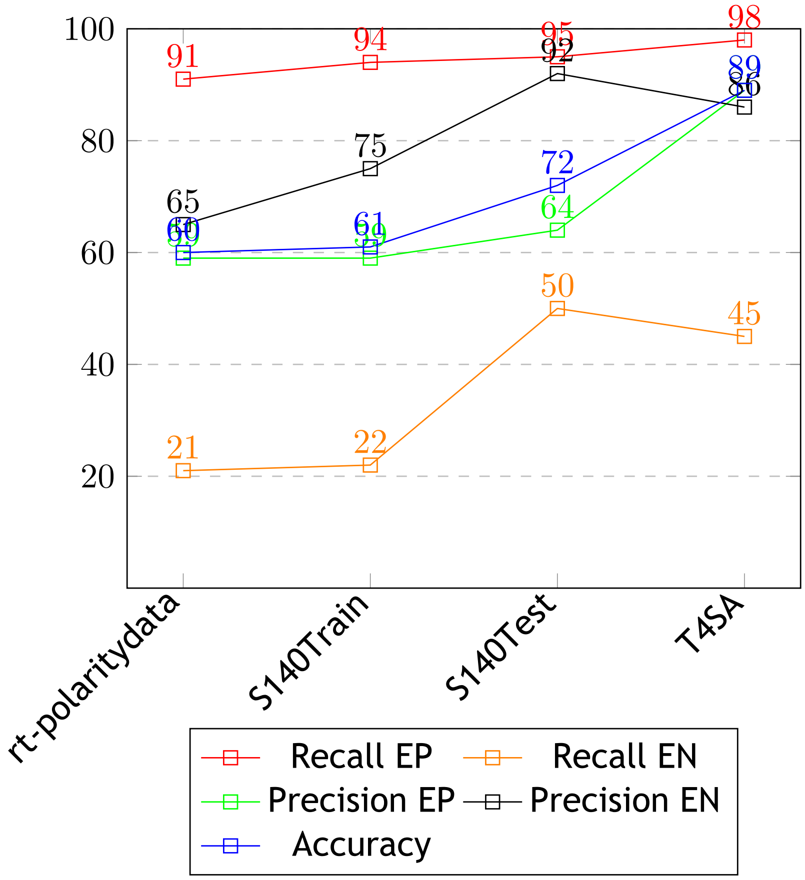

Figure 5.1: Accuracy, Recall and Precision Percentage per Dataset

In table 5.6 and figure 5.1 we have the calculations of recall, precision, accuracy of result in each table. The most notable results are *Recall EP* which are the best results and *Recall EN* which are the worst. The worst results are due to the situations such as the unbalanced ExtremeSentiLex and datasets. The other measures are not so good as the *Recall EP* but are pretty good on all datasets. It reveals that ExtremeSentiLex is working good, but there still some improvement to be done, some of this improvements are suggested in the chapter 6.

5.1.2 Results for ExtremeSentilex with word embeddings

In this subsection we find the results obtain by the execution ESC by using the Extended ExtremeSentiLex with word embeddings as an input, the other input is the dataset. As in the last subsection we have the tables with this results. When we see the results of table 5.7 the first thing to be noticed is the lower amount of *INC* result have decreased, which might not be good. This will happen with all the tests with the different datasets. Also, on both tables 5.7 and 5.8, we see that the number of positive extremes has grown more than the negative extremes detected.

Table 5.7: rt-polarity results for lexicon with word embeddings

	P	N	Total
EP	1235	973	2208
EN	112	198	310
INC	3984	4160	8144
Total	5331	5331	10662

Table 5.8: T4SA results for lexicon with word embeddings

	P	N	Neutral	Total
EP	213780	60375	106935	381090
EN	4627	30072	7900	42599
INC	152934	88603	514731	756268
Total	371341	179050	629566	1179957

Table 5.9: Sentiment140 training results for lexicon with word embeddings

	P	N	Total
EP	125330	91545	216875
EN	5510	18138	23648
INC	669160	690316	1359476
Total	800000	799999	1599999

Table 5.10: Sentiment140 test results for lexicon with word embeddings

	P	N	Neutral	Total
EP	29	18	4	51
EN	1	11	0	12
INC	147	150	137	434
Total	177	179	141	497

In tables 5.9 and 5.10 happen the same that happen on tables 5.7 and 5.8. The amount of *INC* post has decreased, and the number os positive extremes has grown more than the negative extremes.

Now it is also expectable that the result of our measures have changed. And taking into account the number on this tables in a general view the result will decrease, specially accuracy once it is used all the values to calculate the accuracy, also the precision may increase once we have more words on ExtremeSentiLex it will get more false values.

Table 5.11: TurnToIslam and Ansar1 results for lexicon with word embeddings

Datasets	EP	EN	P>N	N>P	INC	Total
TurnToIslam	130658	2821	5018	2579	194252	335328
Ansar1	9725	553	679	496	18039	29492

As mentioned earlier 5.11 has a different objective but we also found the number of *INC* as decreased and in this dataset is where this result of the application of word embeddings is more notable, this parameter has really decreased.

Table 5.12: Indicators of algorithm efficiency for lexicon with word embeddings

	Datasets			
	rt-polarity	Sentiment 140		T4SA
		Training	Test	
Recall$_{EP}$	92%	96%	96%	98%
Recall$_{EN}$	17%	17%	38%	40%
Precision$_{EP}$	56%	58%	62%	86%
Precision$_{EN}$	64%	77%	92%	84%
Accuracy	57%	60%	68%	86%

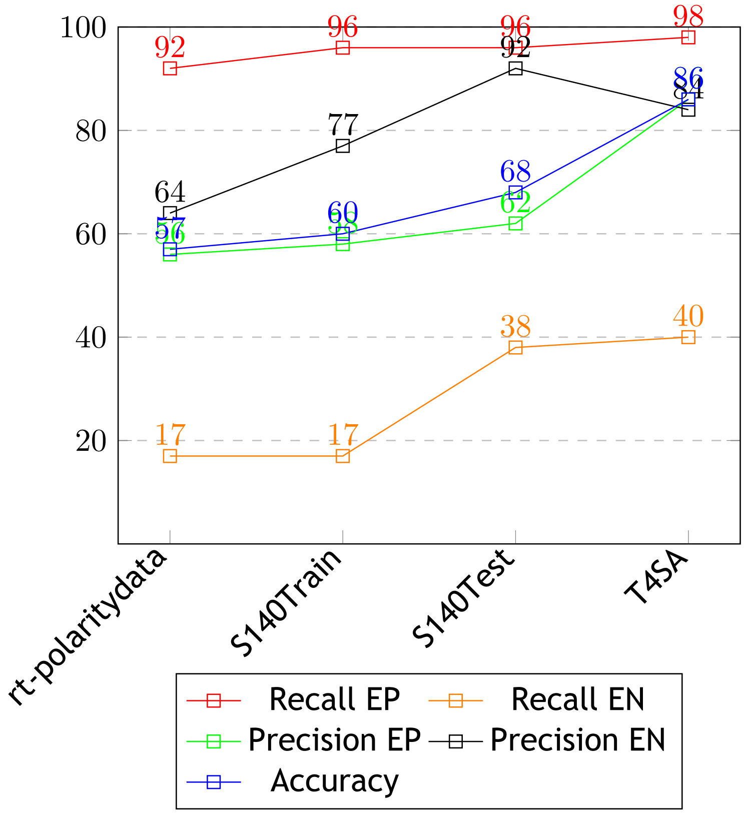

Figure 5.2: Accuracy, Recall and Precision Percentage per Dataset for lexicon with word embeddings

In table 5.12 and figure 5.2 we have the calculations of recall, precision, accuracy of result in each table when using Extended ExtremeSentiLex . The most notable results are, again *Recall EP* which are the best results and *Recall EN* which are the worst.

A comparative study is showed in the section 5.2 but we can already see that in the majority of the datasets the only value to be improved was the recall EP, which was already very high, so it was not a big improvement, the majority of the other values have decreased for all datasets.

5.1.3 Results for Balanced Corpus and Dataset

In this subsection we show a test where we apply three version of ExtremeSentiLex (Original, Extended and Balanced) to a balanced dataset, which is B-T4SA a subset of T4SA.

Table 5.13: Results of B-T4SA with ExtremeSentiLex

	P	N	Neutral	Total
EP	33655	6957	5778	48170
EN	435	5422	256	6113
INC	88772	220483	115048	314303
Total	122862	122862	122862	368586

Table 5.14: Results of B-T4SA with Extended ExtremeSentiLex

	P	N	Neutral	Total
EP	36429	10289	9146	55864
EN	629	6995	314	7938
INC	85804	105578	113401	304784
Total	122862	122862	122862	368586

If we start looking to the table 5.13 we notice immediately a lower difference on the amount of positive and negative posts detected. However, the important part in this subsection is a comparison between all the versions of ExtremeSentiLex. Also in table 5.14 we notice the same differences from the other datasets.

Table 5.15: Results of B-T4SA with balanced ExtremeSentiLex

	P	N	Neutral	Total
EP	4475	2659	4360	11494
EN	444	5436	256	6136
INC	117943	114767	118246	350956
Total	122862	122862	122862	368586

In table 5.15 we see an improvement in the negative part. This is the only test where the amount of negative posts detected as extreme negative posts is higher than the negative posts found as extreme positive posts which at a first view might indicate a that this is the best version of ExtremeSentiLex.

Table 5.16: Indicators of algorithm efficiency for ExtremeSentiLex whit B-T4SA

	ExtremeSentiLex Version		
	Original	Extended	Balanced
Recall$_{EP}$	99%	98%	91%
Recall$_{EN}$	44%	40%	67%
Precision$_{EP}$	83%	78%	63%
Precision$_{EN}$	93%	92%	92%
Accuracy	84%	80%	76%

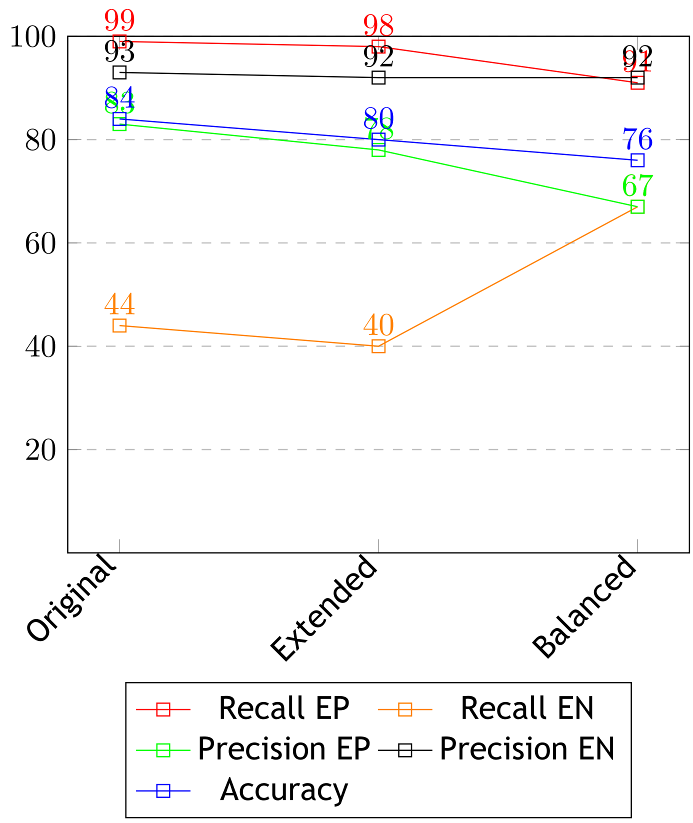

Figure 5.3: Accuracy, Recall and Precision Percentage per version of ExtremeSentiLex applied to B-T4SA

In table 5.16 and figure 5.3 we see something that in despite of the improvement on the negative part when all the inputs are balanced the other values have decreased. In a general view the best results appear when the inputs are a balanced dataset of posts and the original version of ExtremeSentiLex. Furthermore, the results on this dataset are better than in any other dataset we used, this is due to the great quality of the dataset.

5.2 Discussion

After running the experiments, the obtained results are quite encouraging and satisfying, since it is not expected a great amount of extreme posts to be identified as it is not always a case to get extreme sentiments from posts. The most signifying results are in the case of Sentiment 140 test Table 5.6, where none of the calculations is less than 50% and **ESC** performs well on this dataset. Contrary, it is also found some lower values on the negative terms see tables 5.1 and 5.3, the recall for EN and the precision for EP are very low, as shown in Table 5.6. The accuracy for each dataset is either equal to or greater than 60%, which suggests if we flip a coin and choose one side we are more close to choose the right side than the wrong side. Having high precision, we can conclude that choosing the polarity, we are close to getting the correct polarity. In Table 5.2, we try to analyse the result of tweets with images to find extreme sentiments comparing to other datasets with textual posts. We observe that there is no significant change in the results and are not different from other datasets, which means the use of images may not change the tweets intensity.
Next we plot the Table 5.6 results on a graph for better understandability. With our findings among all datasets, T4SA and Sentiment 140 tests are good as all the measures are above 50%.

However, between them, T4SA is better that Sentiment 140 because Sentiment 140 data is very small comparing to T4SA.
We also discovered some major issues and identified a few limitations with our approach during experiments. In case of rt-polarity data (Table 5.1), Sentiment 140 5.3 and T4SA 5.2 datasets, the reason behind low values is due **ESC** not able to distinguish the positive extreme term with negation e.g., *Dems **not Happy** with their nominee*. **ESC** considers happy as an extreme positive term but it is not due to negation. This is something we will handle in future work. Some other issues with TurntoIslam and Anasr1 are language discussion, as few posts are not in English that results in increase inconclusive posts. Problems like long posts with more positive and negative terms also impact the performance of **ESC**, because they are not classified as extremely positive or negative in case of TurntoIslam and Ansar1. To handle it, we classified such posts as inconclusive and categorize them as either P>N and/or N>P, as shown in Table 5.5. Another limitation is the appearance of emojis in posts e.g.,

*Example: If the world ends it will be the most **happy** day of my life :(.*

The **ESC** identifies *happy* as a positive extreme term but the emoji suggests irony.
When we take the first look on the tables with results by applying ExtremeSentiLex with word embeddings we might think it got better once the number of inclusive cases has decreased. But as we will see on comparative graphics below the quality of our work is not related with great quantity. As the number of inconclusive posts on all datasets has decreased we may conclude that by using word embeddings we were able to find more cases of extreme sentiment posts. But it is expected that the number of false extremes might grow. We searched for datasets with content evaluated about extreme sentiment, but we were not able to find any.
On the figures 5.4, 5.5, 5.6, 5.7, and 5.8 we find the comparative studies between the application of ExtremeSentiLex, the last one compares the percentage of $Inconclusive$ cases on each dataset before and after the use of word embeddings to extend our lexicon the others compare the measures of confusion matrix we use to evaluate the quality of our results.

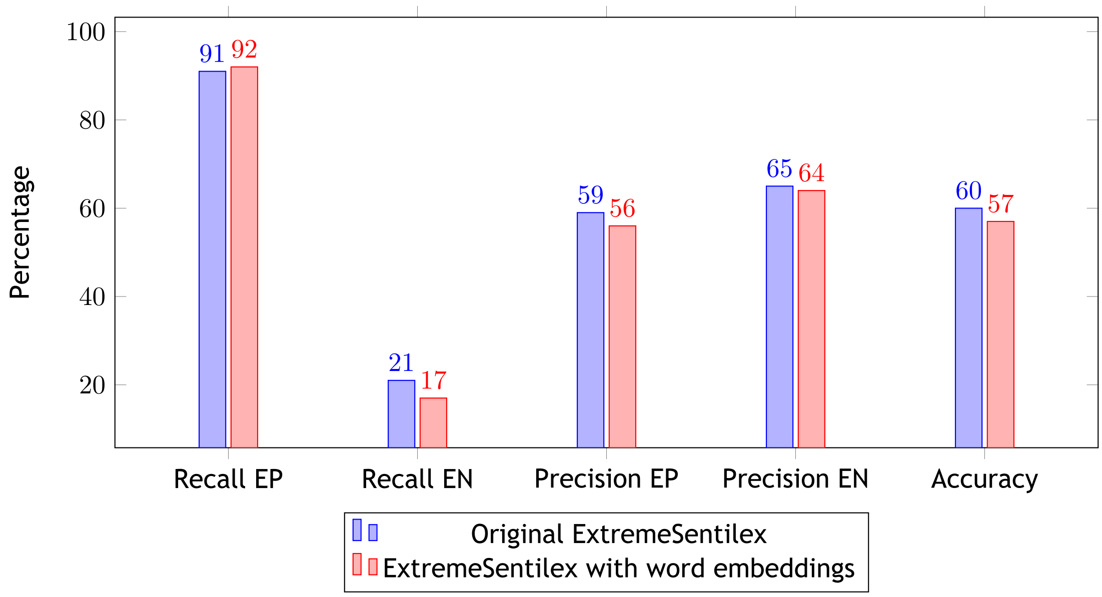

Figure 5.4: Comparison between results of rt-polarity applying ExtremeSentiLex with and without word embeddings

In figure 5.4 we have the result for rt-polarity dataset. As we can see, the only value with a better result is the Recall for Extreme Positive posts, the others decreased which allow us to conclude the result were worst.

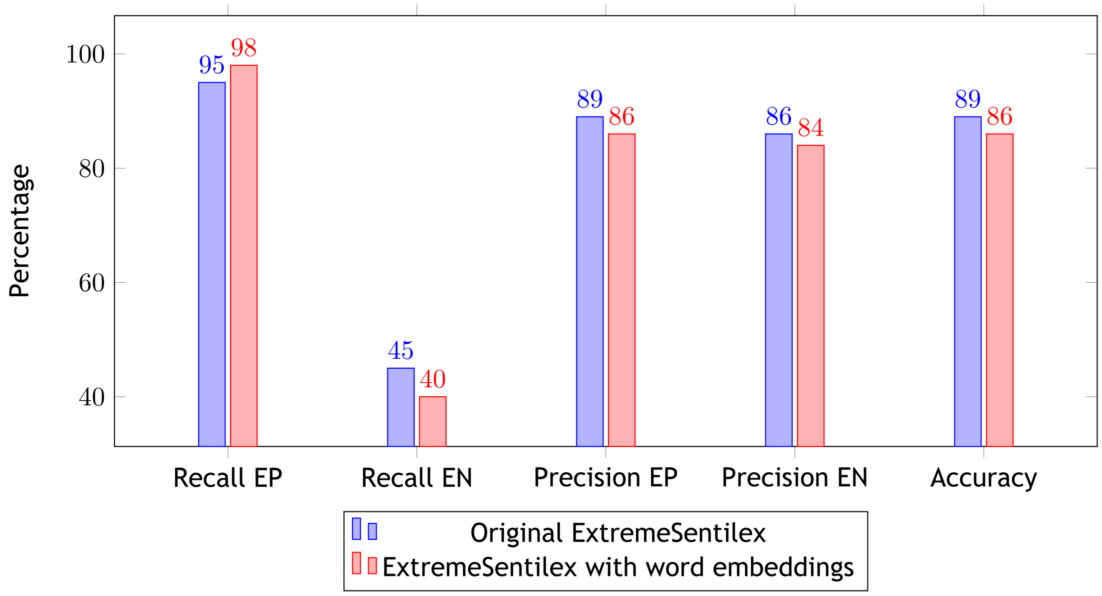

Figure 5.5: Comparison between results of T4SA applying ExtremeSentiLex with and without word embeddings

As expected in T4SA, in figure 5.5, the result is the same. We got better results for Recall of Extreme Sentiment posts, and the other decreased leading us to the same conclusion.

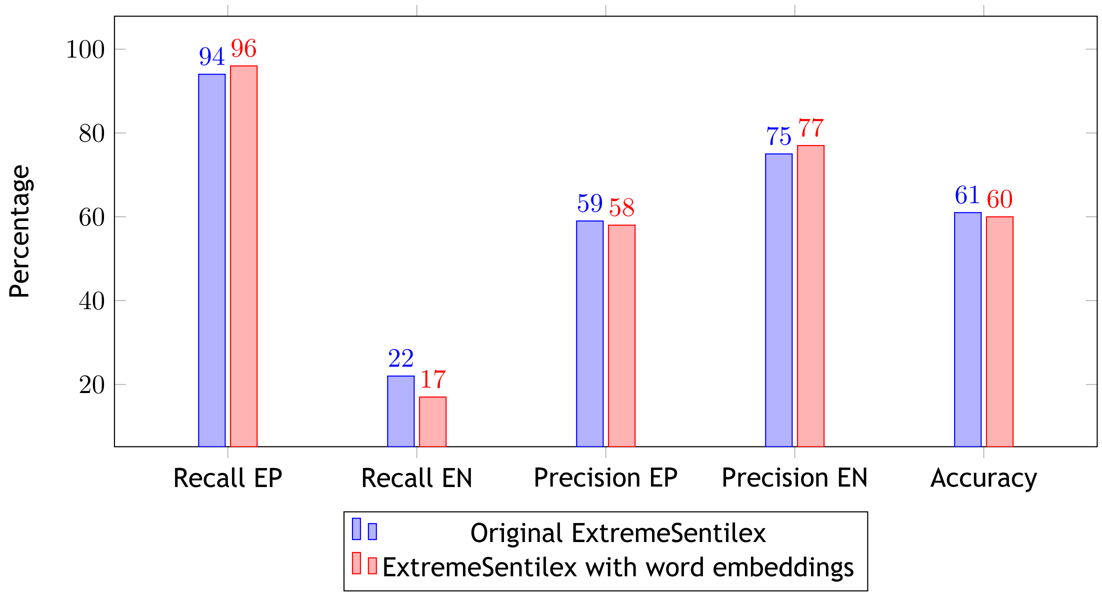

Figure 5.6: Comparison between results of Sentiment140 Training applying ExtremeSentiLex with and without word embeddings

Only in figure 5.6, for Sentiment 140 Training dataset, two values have increased, but the other still decreased, leading to the same conclusion.

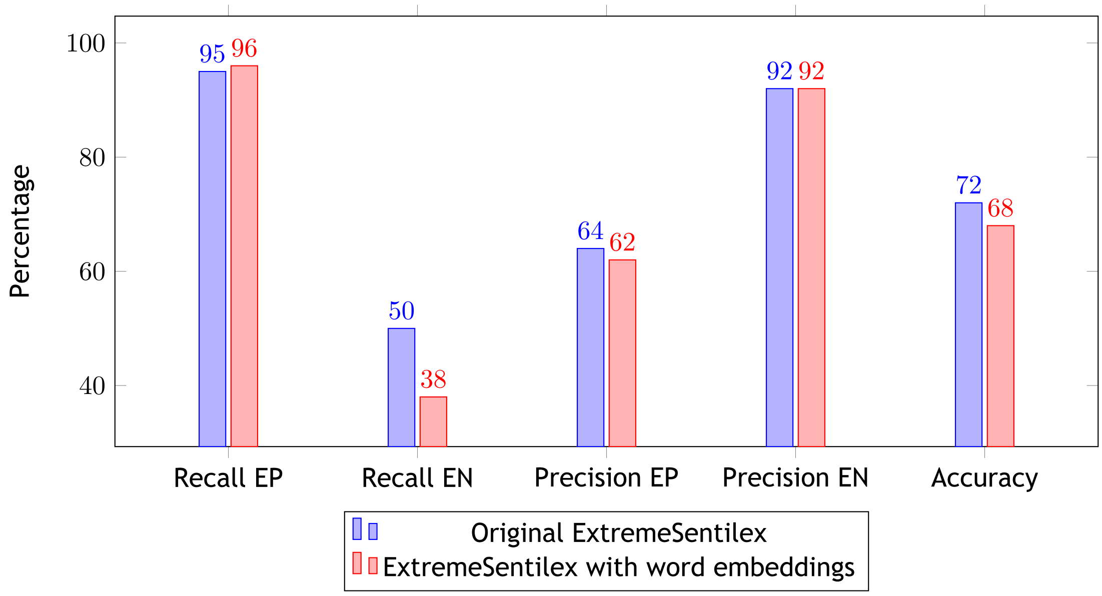

Figure 5.7: Comparison between results of Sentiment140 Test applying ExtremeSentiLex with and without word embeddings

In figure5.7, the Sentiment 140 Test dataset, the Recall for Extreme Positive posts increased and the Precision for Extreme Negative posts is the same, however the other decreased again.

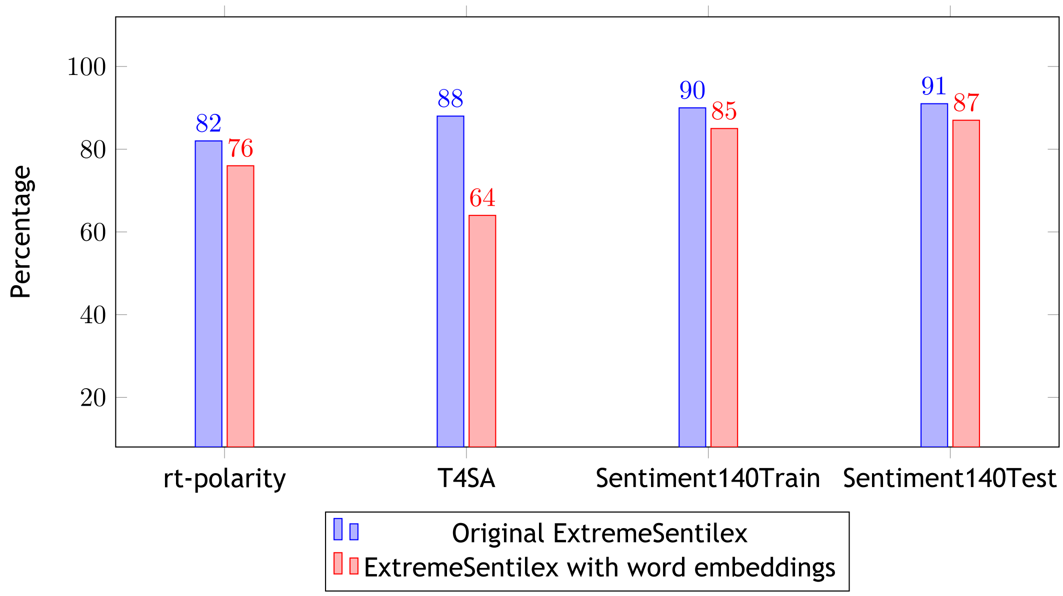

Figure 5.8: Comparison between Inconclusive results in all datasets when applying ExtremeSentiLex with and without word embeddings

In figure 5.8 it is saw a graphic comparing the percentage of $Inconclusive$ posts on each dataset, as we can observe this number decreased very much in all datasets. This leads to a major problem we might be considering extreme post which has nothing to do extremism.

When we applied ExtremeSentiLex extended with word embeddings the results did not coincided whit our hypothesis described in chapter 4. This happened, specially, because once there are more positive terms than negative terms. Being so if we for each term we added five more terms it means that having ten positive terms and 7 negative terms on original lexicon, in the extended lexicon we will have fifty positive terms and thirty-five positive terms so the difference between the amount of positive and negative terms also grown.

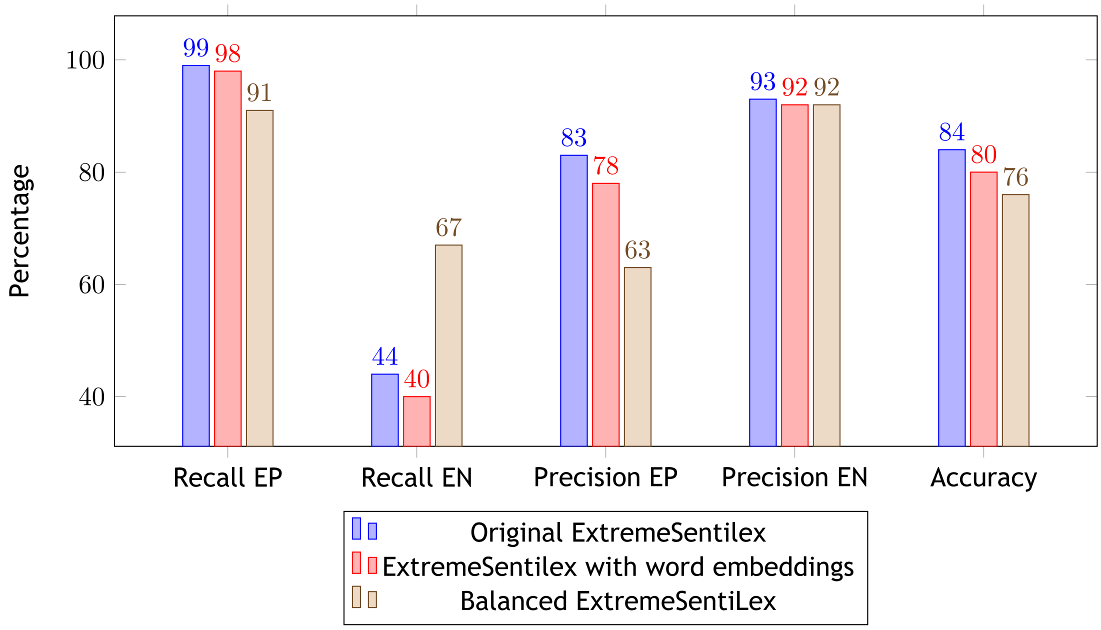

Figure 5.9: Comparison between results of T4SA applying ExtremeSentiLex with word embeddings and the balanced ExtremeSentiLex

In figure 5.9 the results from the experiment of applying a balanced solution based on ExtremeSentiLex is presented, and compared with results by applying the ExtremeSentiLex Original and ExtremeSentilex with the extension with word embeddings. With this results it might be concluded that the values on the extreme negative post recall were improved but the result on the precision of extreme positive posts decreased a lot, on the accuracy as expected the results also decreased.

From this result we open a window where a well balanced lexicon might mean better results, but the filtering used to get a balanced lexicon was not the most convenient. Probably with the help of experts in the linguistic field the results would be even better.

The results have increased due to the values positive and negative terms are not different. This way is possible the number os extreme positive and negative posts are not so different which leads to a more accurate result.

In the ExtremeSentiLex we has a total of 398 terms being 101 negative terms and 297 positive terms. Once we go to the ExtendedExtremeSentiLex has a total of 1121 being 276 of them negative and 845 positive. Once the original lexicon is not balanced the extended version has even a bigger discrepancy between negative and positive terms. It is expected that once we have more positive than negative terms the result on the amount of posts availed as positive will also be bigger, specially due to the discrepancy.

5.3 Conclusion

We demonstrate an approach to detect posts with extreme sentiments on social networks and social media. The main goal of this study is to understand how a lexical resource that consists of extreme sentiments terms can be used to identify extreme sentiments posts on social media and networks. The base of our work relies on two lexical resources, SENTIWORDNET 3.0 [BES10] and

SenticNet5 [CPHK18] by filtering them to obtain only those terms showing extreme sentiments polarity.
To analyse the performance of the system, we apply it on five different datasets and the obtained results are quite satisfactory. Our approach can also be useful for anti-extremism authorities to find issues that involve extreme sentiments e.g., predict terrorism activities or identify radicalism and its causes. Moreover, our methodology is unsupervised and language independent and researchers only need a lexical resource of a certain language to use for their work. The standard lexicon is available on request and can be exploited for research purposes.
Based on the results we obtained in this studies we are capable of conclude that we have a really good lexicon for extremism detection. Once ExtremeSentiLex show better results without the extension using word embeddings. Other points we may have in account is that it might be better if we use a balanced dataset, also the result will depend directly on the type of language used on the post from each dataset.
There is still much work to develop in this area in order to make this work more independent from human evaluation.

Chapter 6

Conclusion and Future Work

In this chapter the final considerations are made regarding the whole process of development of this research work. In addition, approaches are explored that may lead to an improvement not only of the lexicon but also of the process of detecting extreme feelings in social network posts.

6.1 Conclusion

This was a work that took a long time to develop, as it is a work that aims to make various contents available to the scientific community, these contents must be of quality. As we want to provide the files for the scientific community access all the processes for creating and enhancing our lexicon were done in a very minute way and always revised at every step of the term filtering process which ended up using much of the time of the experiment. In addition, the tests to evaluate the quality of our lexicon also took a long time and had to be performed multiple times so that the results were as reliable as possible. Because our goal is the quality of our results because it is these results that will allow us to gauge the quality of our lexicon.
The results presented in section 5.1, are very satisfying once our detection is pretty accurate. When we detect it as positive extreme it has a great probability of being positive post and the same for the negative posts. Furthermore, when we applied the extended lexicon the results have decreased and also the amount of $Inconclusive$ posts have decreased which might mean less accuracy on the detection of extreme posts. This allows us to conclude that ExtremeSentiLex is a really good lexicon to use on the Extreme Sentiment Analysis.
Every process of development this dissertation takes its time, the first process was state-of-the-art development for the topic at hand, this was a dissertation that addressed various areas of knowledge and for this reason, a state-of-the-art was developed. Addressing all the topics we find in this dissertation. After gathering knowledge that allowed us to understand the area in which we will work we began the process of developing our solution to the problem we set out to solve. We then began to create our lexicon always under intensive supervision to ensure the quality of the result obtained. Despite the already high quality of our lexicon, there are ways to still improve it, and the main way would be with the intervention of human experts in the field of extremism in the field of psychology.
This dissertation was an increasingly interesting challenge, in which at each step of the development process several hypotheses of approach to the topic were found, approaches that unfortunately cannot all be explored since the time for the development of a master's dissertation for the sake of time constraints.
In addition to the obvious contributions that this dissertation with archives that are created and accessible to the entire community, it is also a dissertation that, by addressing a very relevant topic today using innovative techniques, leaves an open path to solving a complex problem. and which is proving increasingly important because of the circumstances in which we find society today. The detection of extremism and radicalism in social networks is increasingly a problem

that needs to be addressed to protect lives and avoid clutter in particular communities. The dangers increasingly start in digital and end in reality.
The topic of radicalisation remained unresolved in this dissertation, however, the topic has been properly studied and so we already have the means and the knowledge to solve this topic, the focus to develop is after the detection of a possibly extreme post make a historical analysis user posts and analyse this history, ultimately taking into account the roots of radicalisation, classify the users radicalisation level so that this user is flagged and find other users who might have some connection with the first one.
With this dissertation are open topics that must be addressed for the future for this project to process and be improved, some of these topics are covered in the next section.

6.2 Future Work

As an investigation is being developed is natural that for a specific solution, different paths may appear. So, while we were doing our research many paths come up to our minds. This different path could lead us to different solutions and as we could not cover every topic that appear on our research we needed to choose for the option that were more suitable for our propose.
Below we leave ways to improve our lexicon for the researchers to get better results if they want to create a new version of this lexicon and also for we to keep working on this tool.

6.2.1 Radicalisation

Radicalisation was one of the topics on the table when this research started, but once we started research on extremism we noticed that radicalisation would be hard to develop a system to take that into account. So this is a work to be developed soon, once we are already able to detected post with extreme sentiment we now need to develop a system to analyse the historic of a user which has a post classified as extreme.
As it is described in the state-of-the-art of this investigation, radicalisation is a process of becoming radical, a person may have an act of extremism but this does not define it as a radicalised person. Having a historic of extreme ideals may lead to a process of radicalisation and this people need to be flagged and reported to the competent authorities.
This is suggested the methodology to deal with the issue of radicalisation on digital environment. Nowadays, this is on of the most problematic issues on our world, once the wars on our history happen due to radical ideals, this must be prevented.

6.2.2 Human contributions for the improvement of ExtremeSentiLex

After the developed work was concluded it was clear what is necessary for the improvement of ExtremeSentiLex and it is the contribution from experts on the area of linguistic and psychology in the area of extremism. This kind of knowledge would help in the balancing of ExtremeSentiLex in the amount of positive and negative terms. With a balanced corpus the accuracy of ExtremeSentiLex would be improved.
With the knowledge on this areas it would be possible to analyse ExtremeSentiLex, add extreme terms missing in the lexicon or remove terms we classified as extremes which are not extreme terms. By doing this, the lexicon would be more reliable and it would allow not only a more balanced lexicon but also a richer one.

As we keep seeing nowadays, crossing different knowledge areas is the future for the investigation, much work have been done in independents areas, the time to cross different lines of thinking will allow the humankind to keep evolving. In computer science this kind of work is getting more and more interesting and helpful in the resolution of many problems found in the modern society.

6.2.3 Machine Learning Techniques

On the part of detection of the extreme sentiments on social media post, as far as we know the use of machine learning techniques could improve the time taken for the detection and also the detection accuracy would has more quality. If the system starts to get patterns for extremist language there would be a more effective way of detecting extremism. Once the system would not need the input of a corpus and starts to understand the language in which extremism and radicalism are based, the improvements would be massive.

With this investigation we opened a path for this kind of improvements, once now the community has a lexicon of extreme sentiment terms and also datasets of text evaluated about its extreme sentiment either positive and negative.

Hopefully, in the future there will be work based on this investigation to make our approach even better and new approaches will be created to solve this problem using different machine learning techniques

Bibliography

[AB13] Claudia Aradau and Tobias Blanke. The politics of digital crowds. *Lo Squaderno*, 33:31–38, 2013. 7

[ABDP+08] Rogelio Alonso, T Bjorgo, Donatella Della Porta, Rik Coolsaet, Farhad Khosrokhavar, R Lohker, Magnus Ranstorp, Fernando Reinares, A Schmid, Andrew Silke, et al. Radicalisation processes leading to acts of terrorism. a concise report prepared by the european commission's expert group on violent radicalisation. 2008. 6

[AIL13a] University of Arizona Artificial Intelligence Lab, Management Information Systems Department. Ansar1 forum dataset. University of Arizona Artificial Intelligence Lab, AZSecure-data, Director Hsinchun Chen, 2013. 31

[AIL13b] University of Arizona Artificial Intelligence Lab, Management Information Systems Department. Turn to islam forum dataset. University of Arizona Artificial Intelligence Lab, AZSecure-data, Director Hsinchun Chen, 2013. 31

[BDS14] David Bamman, Chris Dyer, and Noah A Smith. Distributed representations of geographically situated language. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, volume 2, pages 828–834, 2014. 2, 22

[BES10] Stefano Baccianella, Andrea Esuli, and Fabrizio Sebastiani. Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining. In *Lrec*, volume 10, pages 2200–2204, 2010. xii, 27, 28, 45

[BL04] Steven Bird and Edward Loper. Nltk: the natural language toolkit. In *Proceedings of the ACL 2004 on Interactive poster and demonstration sessions*, page 31. Association for Computational Linguistics, 2004. 22

[BNG+11] Hila Becker, Mor Naaman, Luis Gravano, et al. Selecting quality twitter content for events. *ICWSM*, 11, 2011. xi, 1

[CH12] Erik Cambria and Amir Hussain. *Sentic computing: Techniques, tools, and applications*, volume 2. Springer Science & Business Media, 2012. 9

[CPHK18] Erik Cambria, Soujanya Poria, Devamanyu Hazarika, and Kenneth Kwok. Senticnet 5: Discovering conceptual primitives for sentiment analysis by means of context embeddings. In *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018. xii, 27, 28, 46

[Cre18] Xavier Crettiez. « "gilets jaunes": la violence, l'arme des bavards, est aussi celle des silencieux ». *Le Monde*, Dec 2018. Available from: `https://www.lemonde.fr/idees/article/2018/12/04/gilets-jaunes-la-violence-l-arme-des-bavards-est-aussi-celle-des-silencieux_5392285_3232.html?xtmc=gilets_jaunes_extremisme&xtcr=1`. xi, 1

[Cuk10] Kenneth Cukier. *Data, data everywhere: A special report on managing information*. Economist Newspaper, 2010. 11

[DHS12] Richard O Duda, Peter E Hart, and David G Stork. *Pattern classification*. John Wiley & Sons, 2012. 24

[Dia10] Gaël Dias. *Information Digestion*. PhD thesis, Université d'Orléans, 2010. 8

[DM] Sebastifio Pais Gaél Dias and Rumen Moraliyski. Unsupervised and language-independent method to recognize textual entailment by generality. *CLIB 2014 Proceedings*. 21

[DMSR14] Rehab M Duwairi, Raed Marji, Narmeen Sha'ban, and Sally Rushaidat. Sentiment analysis in arabic tweets. In *Information and communication systems (icics), 2014 5th international conference on*, pages 1–6. IEEE, 2014. 9, 10

[DPL12] Donatella Della Porta and Gary LaFree. Guest editorial: Processes of radicalization and de-radicalization. *International Journal of Conflict and Violence (IJCV)*, 6(1):4–10, 2012. 6

[ES06] Andrea Esuli and Fabrizio Sebastiani. Sentiwordnet: A publicly available lexical resource for opinion mining. In *LREC*, volume 6, pages 417–422. Citeseer, 2006. 17

[FAA18] Miriam Fernandez, Moizzah Asif, and Harith Alani. Understanding the roots of radicalisation on twitter. *WebSci '18 Proceedings of the 10th ACM Conference on Web Science*, 2018. 13, 14, 15

[FBSH15] Natalie Friedrich, Timothy D Bowman, Wolfgang G Stock, and Stefanie Haustein. Adapting sentiment analysis for tweets linking to scientific papers. *arXiv preprint arXiv:1507.01967*, 2015. 30

[Fot08] Marcus Foth. *Handbook of research on urban informatics: the practice and promise of the real-time city: the practice and promise of the real-time city*. IGI Global, 2008. 8

[Fuc17] Christian Fuchs. *Social media: A critical introduction*. Sage, 2017. 8

[GBH09] Alec Go, Richa Bhayani, and Lei Huang. Twitter sentiment classification using distant supervision. *CS224N Project Report, Stanford*, 1(12):2009, 2009. 30, 31

[GH15] Amir Gandomi and Murtaza Haider. Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, 35(2):137–144, 2015. 11

[Got16] Gregory Goth. Deep or shallow, nlp is breaking out. *Communications of the ACM*, 59(3):13–16, 2016. 24

[Isk17] Bandar Seri Iskandar. Terrorism detection based on sentiment analysis using machine learning. *Journal of Engineering and Applied Sciences*, 12(3):691–698, 2017. xix, 6, 15, 17, 18

[JDC19] Mark Daniel Jaeger and Myriam Dunn Cavelty. From madness to wisdom: intelligence and the digital crowd. *Intelligence and national security*, 34(3):329–343, 2019. 7

[JSD16] Rajkumar Jagdale, Vishal Shirsat, and Sachin Deshmukh. Sentiment analysis of events from twitter using open source tool. *INTERNATIONAL JOURNAL OF COMPUTER SCIENCE AND MOBILE COMPUTING*, 54:475–485, 05 2016. 23

[KDA+16] Muhammad Taimoor Khan, Mehr Durrani, Armughan Ali, Irum Inayat, Shehzad Khalid, and Kamran Habib Khan. Sentiment analysis and the complex natural language. *Complex Adaptive Systems Modeling*, 4(1):2, 2016. 9

[KDR15] Tom Kenter and Maarten De Rijke. Short text similarity with word embeddings. In *Proceedings of the 24th ACM international on conference on information and knowledge management*, pages 1411–1420. ACM, 2015. 24

[KGB+14] Arie W Kruglanski, Michele J Gelfand, Jocelyn J Bélanger, Anna Sheveland, Malkanthi Hetiarachchi, and Rohan Gunaratna. The psychology of radicalization and deradicalization: How significance quest impacts violent extremism. *Political Psychology*, 35:69–93, 2014. 6

[Kru13] Justine S Krumm. Influence of social media on crowd behavior and the operational environment. Technical report, ARMY COMMAND AND GENERAL STAFF COLLEGE FORT LEAVENWORTH KS SCHOOL OF ?, 2013. xi, 1

[Lim12] Merlyna Lim. Clicks, cabs, and coffee houses: Social media and oppositional movements in egypt, 2004–2011. *Journal of communication*, 62(2):231–248, 2012. 7

[Low17] David Lowe. Prevent strategies: The problems associated in defining extremism: The case of the united kingdom. *Studies in Conflict & Terrorism*, 40(11):917–933, 2017. 5

[LQY+18] Paula Lauren, Guangzhi Qu, Jucheng Yang, Paul Watta, Guang-Bin Huang, and Amaury Lendasse. Generating word embeddings from an extreme learning machine for sentiment analysis and sequence labeling tasks. *Cognitive Computation*, 10(4):625–638, 2018. 24

[MCCD13] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, 2013. 24

[Mer19] Sam Meredith. 'we are at war': Chile's president extends state of emergency after deadly riots, Oct 2019. Available from: `https://www.cnbc.com/2019/10/21/chile-protests-president-extends-state-of-emergency-after-riots.html`. 5

[MHK14] Walaa Medhat, Ahmed Hassan, and Hoda Korashy. Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, 5(4):1093–1113, 2014. xix, 9, 16

[MMS99] Christopher D Manning, Christopher D Manning, and Hinrich Schütze. *Foundations of statistical natural language processing*. MIT press, 1999. 24

[MSC+13] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pages 3111–3119, 2013. 2

[Nas08] Olfa Nasraoui. Web data mining: Exploring hyperlinks, contents, and usage data. *ACM SIGKDD Explorations Newsletter*, 10(2):23–25, 2008. 24

[NEW17] ABC NEWS. *A look at how ISIS is recruiting young Americans through the internet*. 2017. Available from: `https://www.youtube.com/watch?v=4FZCOWWzHQs`. 8

[Nor08] Peter Norvig, Jul 2008. Available from: `http://norvig.com/ngrams`. 20

[Obs15] Observador. *O Estado Islâmico e a Internet: onde e como eles recrutam*. 2015. Available from: https://observador.pt/2015/11/20/o-estado-islamico-e-a-internet-onde-e-como-recruta/. 9

[Par01] W Gerrod Parrott. *Emotions in social psychology: Essential readings*. Psychology Press, 2001. 19

[PK80] R. Plutchik and H. Kellerman. *Theories of emotion*. Emotion, theory, research, and experience. Academic Press, 1980. Available from: https://books.google.pt/books?id=TV99AAAAMAAJ. 1

[PL05] Bo Pang and Lillian Lee. Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. In *Proceedings of the ACL*, 2005. 30, 31

[PL08] Bo Pang and Lillian Lee. Opinion mining and sentiment analysis (foundations and trends (r) in information retrieval), 2008. 28

[PSM14] Jeffrey Pennington, Richard Socher, and Christopher Manning. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1532–1543, 2014. 24

[QR14] Ashequl Qadir and Ellen Riloff. Learning emotion indicators from tweets: Hashtags, hashtag patterns, and phrases. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1203–1209, 2014. xix, 2, 11, 18, 20, 21

[RBL+07] Rajat Raina, Alexis Battle, Honglak Lee, Benjamin Packer, and Andrew Y Ng. Self-taught learning: transfer learning from unlabeled data. In *Proceedings of the 24th international conference on Machine learning*, pages 759–766. ACM, 2007. 2

[Rei19] Catarina Reis. Polícia não vai deixar manifestação da extrema-direita chegar à sede do bloco de esquerda. *Diário de Notícias*, Jan 2019. Available from: https://www.dn.pt/pais/interior/policia-nao-vai-deixar-manifestacao-da-extrema-direita-chegar-a-sede-do-bloco-de-esquerda-10487434.html. xi, 1

[RS06] Deana A Rohlinger and David A Snow. Social psychological perspectives on crowds and social movements. In *Handbook of social psychology*, pages 503–527. Springer, 2006. 7

[SB+12] Ion Smeureanu, Cristian Bucur, et al. Applying supervised opinion mining techniques on online user reviews. *Informatica Economică*, 16(2):81–91, 2012. 30

[Sch13] Alex P Schmid. Radicalisation, de-radicalisation, counter-radicalisation: A conceptual discussion and literature review. *ICCT Research Paper*, 97(1):22, 2013. 6, 13

[Scr07] Roger Scruton. *The Palgrave Macmillan dictionary of political thought*. Springer, 2007. 5

[SD18] K Suresh and R Dillibabu. Designing a machine learning based software risk assessment model using naïve bayes algorithm. *TAGA JOURNAL VOL. 14*, 2018. 22, 23

[SDF18] Ryan Scrivens, Garth Davies, and Richard Frank. Searching for signs of extremism on the web: an introduction to sentiment-based identification of radical authors. *Behavioral sciences of terrorism and political aggression*, 10(1):39–59, 2018. 6

[SGAB15] Raksha Sharma, Mohit Gupta, Astha Agarwal, and Pushpak Bhattacharyya. Adjective intensity and sentiment analysis. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 2520–2526, Lisbon, Portugal, September 2015. Association for Computational Linguistics. Available from: https://www.aclweb.org/anthology/D15-1300. 2

[Shi08] Clay Shirky. *Here comes everybody: The power of organizing without organizations*. Penguin, 2008. 7

[SLH14] Yue Shi, Martha Larson, and Alan Hanjalic. Collaborative filtering beyond the user-item matrix: A survey of the state of the art and future challenges. *ACM Computing Surveys (CSUR)*, 47(1):3, 2014. 14

[sna] Snap bird. Available from: https://snapbird.org/. 18

[SP12] Kate Starbird and Leysia Palen. (how) will the revolution be retweeted?: information diffusion and the 2011 egyptian uprising. In *Proceedings of the acm 2012 conference on computer supported cooperative work*, pages 7–16. ACM, 2012. 7

[Tin17] Kai Ming Ting. *Confusion Matrix*, pages 260–260. Springer US, Boston, MA, 2017. Available from: https://doi.org/10.1007/978-1-4899-7687-1_50. 31, 35

[Toi15] Jurgens du Toit. A first step into machine learning: Building a bayes classifier, Aug 2015. Available from: https://cloudacademy.com/blog/naive-bayes-classifier/. 18

[TWY+14] Duyu Tang, Furu Wei, Nan Yang, Ming Zhou, Ting Liu, and Bing Qin. Learning sentiment-specific word embedding for twitter sentiment classification. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, volume 1, pages 1555–1565, 2014. xii, 2, 13

[VCC+17] Lucia Vadicamo, Fabio Carrara, Andrea Cimino, Stefano Cresci, Felice Dell'Orletta, Fabrizio Falchi, and Maurizio Tesconi. Cross-media learning for image sentiment analysis in the wild. In *2017 IEEE International Conference on Computer Vision Workshops (ICCVW)*, pages 308–317, Oct 2017. 30, 31

[VT13] SM Vohra and JB Teraiya. A comparative study of sentiment analysis techniques. *Journal JIKRCE*, 2(2):313–317, 2013. 24

[Yad16] Danny Yadron. Twitter deletes 125,000 isis accounts and expands anti-terror teams. *The Guardian*, Feb 2016. Available from: https://www.theguardian.com/technology/2016/feb/05/twitter-deletes-isis-accounts-terrorism-online. 6, 15

Appendix A

Appendix

A.1 ExtremeSentiLex Content

grief	-1
tawdry	-1
deceptively	-0.99
lugubriousness	-0.99
shit	-0.99
incendiary	-0.98
infamous	-0.98
lazarus	-0.98
lugubrious	-0.98
notorious	-0.98
unconscionable	-0.98
vitiate	-0.98
wan	-0.98
damage	-0.97
dreaded	-0.97
fingerless	-0.97
latent	-0.97
poverty	-0.97
sickly	-0.97
acne	-0.96
beastliness	-0.96
brutal	-0.96
culpable	-0.96
destructive	-0.96
horrendous	-0.96
indifferent	-0.96
melancholy	-0.96
silly	-0.96
snotty	-0.96
sob	-0.96
agoraphobic	-0.95
blank	-0.95
castigation	-0.95
curdle	-0.95
fallacious	-0.95
filth	-0.95
filthy	-0.95
flutter	-0.95
hardheartedness	-0.95
impeccable	-0.95
rot	-0.95
run-down	-0.95
sordid	-0.95
squalid	-0.95
unemployment	-0.95
unethical	-0.95
venom	-0.95
derogatory	-0.94
disdainful	-0.94
distressed	-0.94

glandular_plague	-0.94
gory	-0.94
incorporeality	-0.94
nonporous	-0.94
preservation	-0.94
profoundly_deaf	-0.94
reserves	-0.94
scornful	-0.94
unconscious	-0.94
unreality	-0.94
anthrax	-0.93
atrociously	-0.93
calamitous	-0.93
closed-door	-0.93
condole_with	-0.93
depigmentation	-0.93
disconcertment	-0.93
erythroderma	-0.93
habit-forming	-0.93
helpless	-0.93
horrible	-0.93
horrifying	-0.93
negative	-0.93
nonmetallic	-0.93
offensive	-0.93
onslaught	-0.93
queer	-0.93
sanguinary	-0.93
sanguineous	-0.93
slaughterous	-0.93
stagnant	-0.93
stupid	-0.93
turgid	-0.93
ungrateful	-0.93
unthinkable	-0.93
badly	-0.92
cattiness	-0.92
hooligan	-0.92
impetuous	-0.92
inclement	-0.92
knockout_punch	-0.92
malevolency	-0.92
naughty	-0.92
nonresident	-0.92
overheat	-0.92
prejudice	-0.92
schizoid	-0.92
short-circuit	-0.92
sterile	-0.92
unlikely	-0.92

unwisely	-0.92
avail	0.87
belle	0.87
boatmanship	0.87
discernible	0.87
evident	0.87
evidently	0.87
expectant	0.87
experimental	0.87
eye_candy	0.87
face-saving	0.87
fascinate	0.87
full-dress	0.87
glowing	0.87
humanitarian	0.87
inspirational	0.87
internal	0.87
mien	0.87
noticeable	0.87
observable	0.87
occur	0.87
outstanding	0.87
overachiever	0.87
pedestal	0.87
perfect	0.87
perspicacious	0.87
philanthropic	0.87
plaudit	0.87
plummy	0.87
popularity	0.87
refreshing	0.87
respectability	0.87
seamanship	0.87
sexless	0.87
sidesplitting	0.87
venerableness	0.87
visible	0.87
well-to-do	0.87
answerable	0.875
absolutely	0.88
adroitness	0.88
alright	0.88
approach	0.88
award	0.88
buildup	0.88
clever	0.88
commission	0.88
cozy_up	0.88
divinity	0.88
enamour	0.88

erogenous	0.88
fecund	0.88
feminize	0.88
fertile	0.88
fertilizable	0.88
fond	0.88
go_over	0.88
good-hearted	0.88
honeymoon	0.88
humanist	0.88
impregnable	0.88
inspired	0.88
intelligence	0.88
invigoration	0.88
lifelike	0.88
lovesome	0.88
lubricant	0.88
luxe	0.88
modest	0.88
mom	0.88
monumental	0.88
overenthusiastic	0.88
pictorial	0.88
play_up	0.88
purist	0.88
ranking	0.88
reliable	0.88
right-hand	0.88
rosy	0.88
season	0.88
shape_up	0.88
shine_up	0.88
sidle_up	0.88
sophisticated	0.88
steadfastly	0.88
superb	0.88
thorough	0.88
toothy	0.88
venal	0.88
vivid	0.88
voluptuary	0.88
artful	0.89
beguiling	0.89
cadge	0.89
charming	0.89
creditable	0.89
dazzle	0.89
depressant	0.89
diacritical	0.89
discerning	0.89

discriminating	0.89
divert	0.89
enthusiasm	0.89
enticing	0.89
fame	0.89
good_weather	0.89
grub	0.89
hairsplitting	0.89
happy	0.89
jocular	0.89
life-or-death	0.89
magazine_rack	0.89
modulated	0.89
mooch	0.89
perfectly	0.89
philistine	0.89
shrewd	0.89
stimulative	0.89
success	0.89
tolerant	0.89
unsociableness	0.89
unwaveringly	0.89
waggish	0.89
whale_louse	0.89
bold	0.9
credible	0.9
desirable	0.9
flourishing	0.9
functionality	0.9
hospitable	0.9
palmy	0.9
permutation	0.9
persevering	0.9
prolific	0.9
prospering	0.9
prosperous	0.9
reliability	0.9
resplendence	0.9
retractile	0.9
salutation	0.9
sensibleness	0.9
sociable	0.9
steadfast	0.9
substantial	0.9
successful	0.9
tactfulness	0.9
thriving	0.9
truth	0.9
venerable	0.9
warrant	0.9

agent	0.91
asserted	0.91
enchant	0.91
firmly	0.91
first-rate	0.91
gracefulness	0.91
healthful	0.91
healthier	0.91
infective_agent	0.91
joke	0.91
liberal	0.91
lusty	0.91
miniaturist	0.91
miraculous	0.91
professed	0.91
protocol	0.91
public_knowledge	0.91
redemptive	0.91
rescue	0.91
responsible	0.91
reverently	0.91
rich	0.91
satisfier	0.91
sincerity	0.91
snow	0.91
sure	0.91
thoroughgoing	0.91
trenchant	0.91
unassertive	0.91
verisimilitude	0.91
amuse	0.92
calculable	0.92
calibre	0.92
coherent	0.92
deign	0.92
everlasting	0.92
fortuitously	0.92
gruntle	0.92
gush	0.92
incisive	0.92
induction	0.92
intelligent	0.92
intricately	0.92
lucky	0.92
measurable	0.92
please	0.92
providential	0.92
selfless	0.92
smart	0.92
smartness	0.92

specular	0.92
stature	0.92
stimulate	0.92
top-notch	0.92
top-quality	0.92
utter	0.92
vouchsafe	0.92
well-off	0.92
wonder	0.92
accept	0.93
applicable	0.93
beauty	0.93
better	0.93
droll	0.93
elegant	0.93
germane	0.93
infatuation	0.93
knowable	0.93
live	0.93
marksmanship	0.93
optimum	0.93
paternalism	0.93
prominent	0.93
quintessential	0.93
racy	0.93
reducible	0.93
slavelike	0.93
tidy	0.93
unsurmountable	0.93
upbeat	0.93
anticipative	0.94
charity	0.94
conspicuous	0.94
disinvolve	0.94
fundamentally	0.94
gravity	0.94
involution	0.94
irenic	0.94
kind	0.94
music	0.94
non-negotiable	0.94
obsequious	0.94
per_se	0.94
resoluteness	0.94
satisfy	0.94
servile	0.94
single-mindedness	0.94
unambitious	0.94
acknowledgeable	0.95
actively	0.95

cognac	0.95
compliant	0.95
dedication	0.95
edified	0.95
enjoyment	0.95
grandness	0.95
idealized	0.95
identifiable	0.95
judiciousness	0.95
moralist	0.95
possess	0.95
probity	0.95
purposive	0.95
recognizable	0.95
rehabilitate	0.95
specifiable	0.95
take_stock	0.95
compassionate	0.96
compliment	0.96
gentle	0.96
indorse	0.96
maintainable	0.96
perspicuous	0.96
pluralize	0.96
respectful	0.96
reusable	0.96
usability	0.96
dignity	0.97
gift_wrap	0.97
mama	0.97
mummy	0.97
sumptuous	0.97
chirpiness	0.98
firm	0.98
riveting	0.98
technical	0.98
clean	0.99
cleanable	0.99
etiquette	0.99
islamism	0.99
renovation	0.99
trustworthiness	0.99
estimable	1
first-class	1
happiness	1
mind-blowing	1
sensational	1
splendid	1