Twitter Gender Classification using Convolutional Neural Network

Author

Asfandyar Nasim Khan

00000117883

MS-15(CSE)

Thesis Supervisor

Dr. Usman Akram

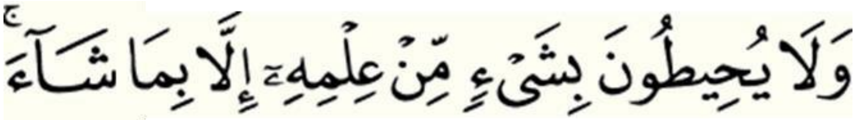
Department of Computer and Software Engineering,

College of Electrical and Mechanical Engineering,

National University of Sciences and Technology, Islamabad.

August 2019

In the name of God, most Gracious, most Compassionate



وَلَا يُحِيطُونَ بِشَيْءٍ مِّنْ عِلْمِهِۦٓ إِلَّا بِمَا شَآءَ ۚ

And they can’t encompass anything from His knowledge, but to extend He wills [2:255]

TWITTER GENDER CLASSIFICATION USING CONVOLUTIONAL NEURAL NETWORK

Author
ASFANDYAR NASIM KHAN
00000117883

A thesis submitted in partial fulfillment of the requirements for the degree of
MS Computer Software Engineering

Thesis Supervisor:
DR. USMAN AKRAM

DEPARTMENT OF COMPUTER AND SOFTWARE ENGINEERING, COLLEGE OF ELECTRICAL & MECHANICAL ENGINEERING, NATIONAL UNIVERSITY OF SCIENCES AND TECHNOLOGY,
ISLAMABAD.
AUGUST 2019

Language Correctness Certificate

This thesis has been read by an expert of the English language and is found free of typing, syntax, semantic, grammatical and spelling mistakes. The stated is also according to the format given by the university.

Signature of Student
Asfandyar Nasim Khan
Registration Number
00000117883

Signature of Supervisor
Dr. Usman Akram

Declaration

I certify that this research work titled "*Twitter Gender Classification using Convolutional Neural Network"* is my own work. The work has not been presented elsewhere for assessment. Material derived from other sources has been properly acknowledged/referred.

Signature of Student
Asfandyar Nasim Khan
MS-15(CSE)
00000117883

Copyright Statement

Acknowledgements

I am extremely thankful to my ALLAH, The Almighty, for bestowing countless blessings upon me and vesting in me patience and ability to complete this manuscript. It is merely because of His kindness and assistance which enabled me to endure and succeed. I am very grateful to my supervisor **Dr. Usman Akram** for his guidance and efforts. Throughout this journey, he has always been there to guide me and boost my confidence. He has been a source of great knowledge and his expertise made this possible. Furthermore, I want to extend my heartiest gratitude to **Dr. Arslan Shaukat** and **Dr. Farhan Hussain** for their availability, guidance, support, and valuable suggestions. I want to thank my parents and wife for their unconditional love, support and prayers. I am grateful to my siblings and my friends for motivating and encouraging me.

Dedicated to my adorable daughter

Abstract

Lately, social media has become one of the favorite topics of fields like Data Science, Machine Learning, Data Mining, Big Data, and Natural Language Processing. This is due to the fact that data is abundantly present on social media platforms. These platforms include Facebook, Twitter, Instagram, and Flickr, etc. Gaining some insight into user data can be of great use when it comes to tailored campaigns like advertisements, or political campaigns. Gender prediction also possesses special significance when it comes to other domains where the identification of an organization is important. For example, emergency management and on other occasions where classifying between male or female is critical for instance in campaigns that are directed towards the issues or awareness of gender-based ferocity.

Taking the significance of gender prediction into consideration, this research tries to assess and evaluate a readily presented approach to automatically detect the gender of the users based on provided tweets. This can be helpful in targeting a specific gender group for advertisements or for social media campaigns. As social media campaigns are really helpful in educating a wide range of people with different backgrounds and geographical locations. Convolutional Neural Network or more commonly known as CNN has been used for this categorization. CNN is mostly used for image classification but it is also helpful in text classification. CNN has been made use of for classifying user's gender by considering the texts from their tweets.

CrowdFlower dataset has been used in this thesis. After preprocessing the user tweets are inputted to the CNN where the embedding layer receives polished tweets. It is quite usual to use forward or backward propagation with neural networks but here Adaptive moment estimation technique has been used for weight optimization.

The mean accuracy that has been achieved by the proposed system is 97%. The stated figure is close to 100 percent and thus the proposed system can be used to form an automated prediction system and can be made use of for numerous purposes including tailored advertisements.

In the future, different combinations of weight optimization and loss functions can be used to further improve the performance of the proposed system.

Keywords: *CNN, Social Media, Gender Prediction, NLP, and ML.*

Table of Contents

List of Figures

List of Tables

CHAPTER 1

INTRODUCTION

As far as the area and problem of gender classification are concerned, the interest of researchers has been rising, particularly in marketing domains and social media. A major source of this rising interest is due to the fact that sources of user-related information have been increasing abruptly, these sources include short tweets and comments on blog posts, etc. Systems or techniques that have been readily proposed and which are functional, make use of features for instance parts of speech n-grams, and word classes, etc. for training and classification. However, deep learning has not been used as such, or its models have not been extended to different sources for instance from media sphere to blogging [1].

Numerous papers have been proposed for studying this topic i.e. gender classification in texts [2] [3] [4]. All these papers have treated the stated topic or problem as a classic example of machine learning, and most of them have used non-linear classifiers using tailored word-based attributes or features on various sources of texts. [5] has made use of the British National Corpus in order to study gender and text associations and relationships in writing (formal). The stated study has been able to discover that there exists a clear difference in the style of writing of males and females. Whereas, [2] discovers the effectiveness of tweets for identifying the gender of the author. The approach put forth by them makes use of n-grams concatenated with the information of the author profile for predicting their gender. The stated research has achieved an accuracy of 77% by making use of text only, and when they included all the features their accuracy went beyond 90%. The present state of the art which makes use of pure textual features is actually presented by Liu and Mukherjee [3], who consider content words, dictionary-based content analysis, and parts of speech tags for blog posts.

The abrupt increase in the occurrence of online social media for the exchange of informal information has empowered statistical modeling on a large scale of the association between language style and other social attributes, for instance, geographical origin, race, age, and gender, etc. More often than not there is an implicit presumption that linguistic options are related to essential and immutable groups of people, despite the fact that the objective of research might be to understand stylistic variance or to learn about the predictive models of “latent attributes”. Without any doubt, it is possible to show or reveal sturdy correlations among such categories and language, which enable predictive modeling which is delightfully accurate. However, this causes a misleading and oversimplified image of how language may reveal personal identity [6].

As far as Social Media is concerned, it is a group of applications and services developed on the technological fundamentals of Web 2.0, which empowers or enables the creation and exchange of contents (collaborative), such content is also called user-generated content or UGC. More precisely, Web 2.0 can be termed as a novel methodology which is used by programmers or developers and also by final users in order to make use of the WWW. With the invent of Web 2.0 applications and contents are no longer developed or published merely by individuals but in a manner that is collaborative.

Various methodologies of text, image, video and sound processing can be made use of and applied for finding tendencies, patterns and provide quantitative and qualitative measures for utilization in many areas. These areas include government, economy, recommendation systems, politics and controlling rumors. Despite the growth that has been witnessed over the recent years, an increase in interest in the programmed characterization of social media users or profiles based on the content they generate/create and share is observed.

1.1Overview

Millions of users around the globe consider Twitter, a successful microblogging social media website/platform, as an integral part of their daily life. Besides the informal communication with friends, acquaintances, and family, these services can be used as a service of recommendation as well as sources of real-time news and venues of content sharing.

The experience of a user with a service of microblogging can be improved significantly if the description or information or personal interest or demographic attributes about a specific user as well as general users are available. The stated information can be used for recommending personalized users to follow, some posts to go through, moreover, topics and events of interest can also be highlighted or emphasized for a specific group of people or communities.

Most of the social media platforms or microblogging services contain information of user-profiles such as a summary of interests, name, age, and location, etc. Though there is a possibility that the stated details might not be complete (due to the fact that the users chose not to put it) or misguiding (the user put a wonderland as their location). Moreover, other relevant information, for instance, implicit and explicit interests/likings or political liking/disliking are not listed.

However, despite the fact that such information might be fake or imaginary but there are some trends that exist and can be used to predict the gender of the user.

1.2Background and Motivation

The rapid and abrupt growth of real-time services of microblogging, for example, Facebook and Twitter have caused a mindboggling increase in the efforts put forth to make use of the contents of social media. For instance, major players of web technology, for instance, Google, Yahoo, and Bing are now incorporating microblogging posts and are trending its analysis in their own results. Besides the fact that they are making use of social information in aggregation with residing searching and retrieving models, notable efforts have been carried out for the development of novel applications for example post and user recommendation services.

What motivates this thesis is the fact that the problem of most interest and attention in this regard is the automatic classification of users or profiles, moreover, demographic attributes such as origin, ethnicity, gender, and age, can be mined to achieve the stated objective. Also, interests and liking or disliking such as politics, soccer or a soccer team, a TV series, etc. can also be used for carrying out predictions.

1.3Objective and Contributions

The objective of this thesis is to assess the performance of a readily proposed system and analyze its performance measures such as accuracy when a different dataset is fed to it.

This manuscript contributes to the literature of techniques applied to Social Media datasets as far as classification is concerned in a way that it has selected a different dataset and has applied some preprocessing techniques in order to ready it for training and validating purposes. It then has conducted some experimentations with the parameters and hyperparameters of the proposed model in order to improve the performance of the presented model. And the results reveal that the performance of the proposed system has actually enhanced.

1.4Outline

This is the first chapter of the thesis, and apart from this, there are five more chapters. The second chapter contains the literature review i.e. previously conducted related scientific work. A literature review is quite significant as far as conceiving and perceiving the details of a particular topic. The third chapter provides the proposed methodology presented by this thesis, which explains everything not only figuratively but descriptively in quite some detail. The fourth chapter has the implementation of the proposed methodology and explains the complete architecture of the convolutional neural network, it also provides details about the system on which the methodology has been implemented, and the tools that have been used for implementing the methodology. Whereas the fifth chapter is about the results i.e. the whereabouts of its performance or assessment and comparison with another state-of-the-art techniques. The same chapter also presents an analysis of the proposed technique. Last but not least, the sixth chapter provides the reader with the conclusion and future work of this manuscript.

This chapter is used to provide the reader with the introduction of the problem domain of the thesis. Social media microblogging has lately become quite a thing and the data created and shared on social media is phenomenal. This chapter contains an overview of the thesis. It also presents the background and motivation, and objectives and contribution. It outlines the thesis in order to let the reader know about the contents of the manuscript.

CHAPTER 2

LITERATURE REVIEW

Acquiring in-depth information regarding previously conducted research and literature is of quite some significance and can be termed as a necessary feature of any project be the project an academic one or an industrial one. What increases the significance of the literature review is the fact that it should be capable of creating a concrete base for advancing information. Literature review plays a vital role as far as the facilitation of theory development is concerned, not only that but it also shuts areas where an excessive amount of research readily presides and discovers fields where the further need of research is felt.

In this chapter of the manuscript, a survey of existing machine learning and deep learning techniques that have been applied on Twitter datasets for the identification or classification of gender has been conducted. But before we proceed toward that discussion, we need to illustrate certain terms and areas. This chapter begins with the discussion of Social Media, its emergence, the amount of data generated on Social Media, what benefits can be achieved by making use of that data, and which machine learning, deep learning, and natural language processing techniques have been applied to make use of the data generated on Social Media platforms. Lastly, a table that contains the recent studies conducted on Twitter gender classification has been provided.

2.1Social Media

Conventionally, the main use of the internet had been expending content. A user would make use of the internet for reading, watching and buying or selling services or products. But with the passage of time, consumers started making using other platforms, for instance, wikis, sharing sites, social networking sites, blogs, and others in order to access, modify, create, share and converse over the contents of the internet. This can be used to represent the phenomenon of social media, which now possesses a significant power to influence the reputation of a firm, its sales and even its survival [8].

Web-based and mobile applications have been employed by social media for creating platforms that are highly interactive with which communities and individuals share, discuss, create, co-create, modify and discuss content that is generated by users. If we look at the extraordinary exposure of social media, we would be able to realize that this is an era of a novel communication landscape. According to a report, the New York Times hired an editor of social media where a webinar was offered on how the church can make use of social media. Even the smallest of brands, for example, a milk brand called Northwest Organic Valley has started displaying 'find, friend and follow us' on their milk cartons. The fact that social media actually commenced with a social network called Sixdegrees in 1997, is not known to many. Just like today's social sites, Sixdegrees, allowed users to make profiles, enlist friends, and add friends of their friends to their list.

Today, there is a diverse and rich ecology of social media sites, this varies when it comes to functionality and scope. For instance, some of the sites are for the general masses like Facebook and Instagram. Then there are other sites for example LinkedIn, which is termed as a professional network. As a matter of fact, Facebook, for starters was only meant as a private network for the students of Harvard University. Then there are media sharing sites, for instance, Flickr, MySpace and of course YouTube, their main concentration is on sharing videos. In the 90's a trend of weblogs (blogs) started, which became extremely popular, due to the fact that their creation and maintenance was quite easy. The authors of these weblogs ranged from laymen to celebrities and professional writers. As a matter of fact, in the early 2000's 'blogosphere' had become a source of over 100 million blogs, which had become a significant source of public opinions. And other techies had started coming up with search engines, for example, Technorati which could be used for blog searching. Likewise, there now are sites like Digg, Reddit, and Delicious which can be used for ranking sites by voting their content. And just like Twitter, microblogging came into being which focused on updates which were real-time [8].

Today, Social Media has become a source of top agenda for numerous executives of the business. Not only consultants but decision-makers try to recognize trends and ways in which their firms can profitably make use of applications like YouTube, Twitter, Instagram, and Facebook. Despite the fact that people have advanced so much, there is still limited understanding of the term 'Social Media'. This subsection is intended to clarify the concept of social media. An explanation of the concept of Social Media has been provided at first and then the difference between other concepts like Web 2.0 and User Generated Content is provided [9].

As stated earlier, the concept of Social Media is not groundbreaking. Nonetheless, a sense of confusion seems to be present not only among academic researchers but managers regarding what should be under the umbrella of this term. And how does it differ from something like User Generated Content and Web 2.0 which seem to be interchangeable? That is why it does make sense that a step should be taken back in order to give some insight related to Social Media and its origin.

By the year 1979, Jim Ellis and Tom Truscott from Duke University made a worldwide discussion system called Usenet. This system allowed users of the Internet to post public messages. It can also be claimed that the actual era of Social Media started when Bruce and Susan Abelson developed ‘Open Diary’, which could be termed as a site for social networking that brought online diary writers together and formed a community. This was when the term ‘weblog’ was first coined which was then shortened to become ‘blog’. Later, the Internet gained speed which expanded the commonness of the idea, and sites of social networking such as MySpace was brought forth and a year later Facebook was developed. This is when the term ‘Social Media’ was first coined and subsidized to the popularity and eminence that it has today [9].

Despite the fact that an idea can be formed by looking at the mentioned applications, it is necessary to draw a line to distinguish the two related or often confused concepts which are User Generated Content and Web 2.0. Web 2.0 was first introduced in 2004 in order to illustrate a novel way in which both developers and users started utilizing the World Wide Web. That the WWW is a platform where applications and content are not made and published by individuals anymore but rather are modified continuously by all the users in a collaborative and participatory fashion. Whereas applications for instance Britannica Online and Encyclopedia and the concept of content publication belonged to the period of Web 1.0, these are not taken over by Blogs, Wikis and other collaborative projects in the later version of the Web i.e. Web 2.0. In order to function, Web 2.0 has to adopt a few functionalities despite the fact that it does not refer to any particular update of WWW. These include Adobe Flash (a famous technique introduced for the addition of animation, audio/video streams, and interactivity to webpages), Really Simple Syndication (or more commonly known as RSS which is a family of web feed formats utilized for publishing frequently updated content like blog entries and news headlines), and Asynchronous JavaScript (or more commonly known as AJAX, which is a method for asynchronously retrieving data from web servers, and allowing the web content to update without meddling the outlook and behavior of the page). It can be stated that Web 2.0 provided a platform which enabled the evolution of the Social Media [9].

Web 2.0 can be used to represent technological and ideological foundations, UGC or User Generated Content can be perceived as the combination of all the possibilities in which people use Social Media. UGC gained fame around the year 2005, it is commonly implemented in order to illustrate the different shapes and forms of media content which are available publicly and are created by users that are laymen. According to the OECD or the Organization for Economic Cooperation and Development, UGC has to meet three fundamental requirements to be perceived as such; it has to get published on either a publicly reachable site or on a site that is used for social networking that is accessible to a chosen group of users, it has to depict a particular amount of creativity, and lastly, it has to create other than professional practices and routines. It excludes content communicated in emails or instant messages in order to meet the first condition, as far as the second condition is concerned, a duplication of readily existing content (copy of a newspaper excerpt without comments or modifications) will not qualify, and thirdly, all content which is designed bearing in mind a commercial market does not qualify. As a matter of fact, UGC was present before the concept of Web 2.0 was coined in, however, as described above, the collection of technological economic, and social drivers causes an essential difference in UGC than the one present in the 80s. After providing these basic clarifications regarding Web 2.0 and UGC, it is now clear to provide a detailed illustration of what is meant by Social Media. Social Media can be termed as a collection of applications that are Internet-based and that build on the technological and ideological foundations of Web 2.0, and they simultaneously create and exchange UGC [2].

There are numerous types of Social Media which meet the provided definition, and which need to be categorized further. Wikipedia, Facebook, Twitter, YouTube, and Second Life all fall in the larger group, and there does not exist a systematic procedure of dividing these Social Media platforms into sub-categories. As a matter of fact, novel sites do appear in cyberspace on a daily basis, and hence those sites must also be taken into consideration if a classification scheme has to be made [9].

2.1.1 Data Generated

The amount of data that is generated on Social Media is indeed mind-blowing. Based on our current pace, approximately 2.5 quintillion bytes of data is generated every day. And this pace is only going to increase now that the Internet of Things has been introduced. Merely in the last two years, 90 percent of all the data in the world has been created. Yes, it needs to be reread 90 percent of the data. It is quite unbelievable and that is why some of the stats, regarding some of the ways this humongous amount of data is created, have been enlisted here [10].

2.1.1.1 Internet

Today, we have so much data and not like the previous generations, we have it at our fingertips. And the moment we turn to the search engines for information and answers, we start adding to the stockpile of the data. More than half of the web searches are performed on cell phones now. Over 3.7 billion humans have access to and use the Internet, it has grown about 7.5 percent as compared to 2016. An average per second processes of Google is around 40,000 that is around 3.5 billion searches every day. As a matter of fact, Google alone is used for 77% of the searches, but it will be careless to exclude other search engines which also contribute to the data generation that occur on daily basis. Around the globe, the total number per day is 5 billion [10].

2.1.1.2 Social Media

The affair we are having with Social Media is not only fueling data generation but is also resulting in a colossal amount of data generation itself. According to a report brought forth by Domo Data Never Sleeps, the following numbers are generated every 60 seconds of a day.

- **Snapchat**: 527,760 photos
- **LinkedIn**: 120 new professionals join
- **YouTube**: 4,146,600 users watch videos
- **Twitter**: 456,000 tweets
- **Instagram**: 46,740 photos posted

Facebook has over 2 billion active users, and with those figures, it still is the largest platform of social media. That is actually over a quarter of the world’s population. Here are some more mindboggling facts about Facebook.

- Over 1.5 billion people are active every day on Facebook

- 307 million European people are on Facebook
- In every second, 5 new profiles are made on Facebook
- Every day over 300 million pictures are uploaded to Facebook
- In every 60 seconds 293,000 statuses are uploaded, and in the same duration 510,000 comments are posted.

Despite the fact that Facebook has been the largest social network for a while, Instagram which is also owned by Facebook has been growing impressively. Our data deluge is influenced in the following manner by the photo-sharing platform.

- Out of the 600 million Instagram users, 400 million are active every day
- Around 100 million pictures and videos are shared every day on Instagram
- And the “stories” features are used by 100 million users every day.

2.1.1.3 Communication

A trail of data is left behind whilst we make use of our favorite methods of communication, that includes everything from sending texts to emails. The stats of data regarding our communications is even more intriguing. Every minute we send out,

- 16 million text messages
- 990,000 swipes on Tinder
- 156 million emails (by around 2.9 billion users of email)
- 15,000 GIFs on Facebook messenger
- 103,447,520 spam emails
- 154,200 Skype calls

A more detailed and graphical presentation of some of the stats has been presented in Figure 2-1.

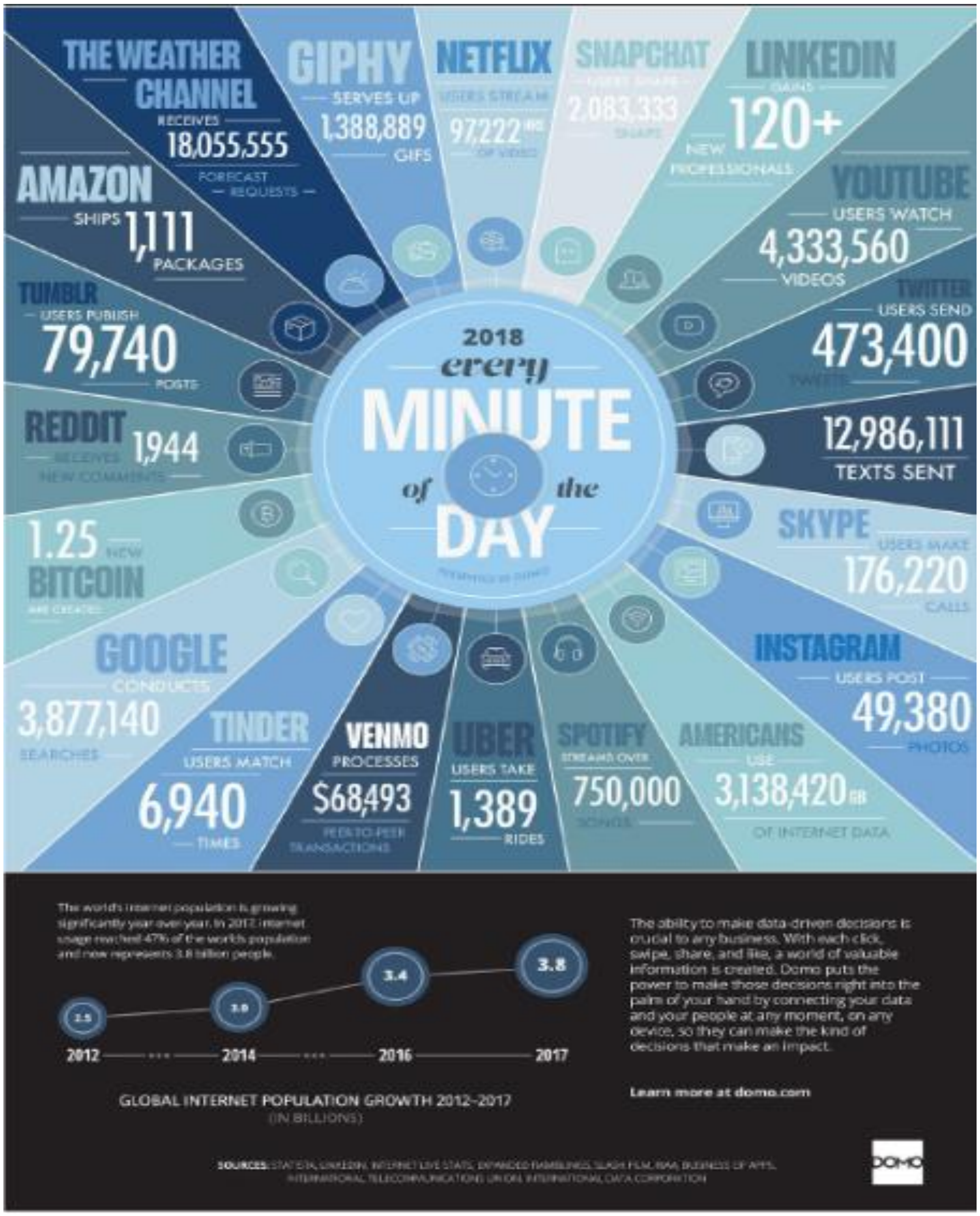
2.1.1.4 Digital Photos

As our phones have become smartphones now, and they do come with exemplary cameras, every other person has become a photographer and the fact that trillions of photos are stored is its proof. Due to the fact that this does not seem to slow down, the 4.7 trillion photos stored in 2017 would now have grown.

2.1.1.5 Services

The economy is now driven by this new platform, and here are some extremely intriguing stats which are presented by service providers and businesses. These numbers are generated every 60 seconds,

- Weather channels get 18,055,556 requests of forecast
- Addition of 13 songs to Spotify



The world's internet population is growing significantly year over year. In 2017 internet usage reached 47% of the world's population and now represents 3.8 billion people.

The ability to make data-driven decisions is crucial to any business. With each click, swipe, share, and like, a world of valuable information is created. Domo puts the power to make those decisions right into the palm of your hand by connecting your data and your people at any moment, on any device, so they can make the kind of decisions that make an impact.

Learn more at domo.com

SOURCES: STATISTA, LINKEDIN, INTERNET LIVE STATS, EXPANDED RAMBLINGS, SLASH FILM, RAA, BUSINESS OF APPS, INTERNATIONAL TELECOMMUNICATIONS UNION, INTERNATIONAL DATA CORPORATION

Figure 2- 1: How much data is generated every 60 seconds? [11]

- Peer to peer transactions of $51,892 on Venmo

- 45,788 trips on Uber
- 600 edits to Wikipedia (new page edits)

2.1.1.6 Internet of Things

The IoT i.e. the Internet of Things which is used to connect smart devices which interact with each other and with us, and in the meanwhile, it is used to collect data of all sorts. In 2006 it used to connect 2 billion devices and the number is projected to go beyond 200 billion devices by 2020. Voice search, just a single device based on IoT, let us have a look at some of its stats,

- Over 30 million voice-first devices
- Every month voice control is used by 8 million people
- Queries related to voice search in Google for 2008 increased 35 times in 2016

Now that we are aware of the fact that data is generated in an extremely high rate in a span of seconds, we can now imagine all the ways data is generated collectively every day. What is of more emphasis is that this speed and fashion is increasing exponentially with every passing day [10].

2.1.2 Making Use of the Data Generated

Most of us, and if put as a figure over 99% of us stop at the mere production and creation of data. While the least of us that is a fraction of 1% make use of the data that is not only generated but is available to be made use of. Numerous techniques, such as machine learning, deep learning, and natural language processing, etc. are used to make use of data. The usage includes proposing an advertisement for a particular audience to life-saving events.

2.2Machine Learning

Learning can be termed as the process of developing some model after the information is retrieved from data, as far as machine learning is concerned it can be illustrated as a computational procedure rather a complex computational procedure which is used to automatically recognize patterns and make intelligent decisions or predictions based on trained samples of data. Machine learning can be termed as something which falls under the arc or umbrella of artificial intelligence. Machine learning is the learning capability of a machine, this learning process is carried out when a decent amount of data is fed to it then the machine is trained on the provided data. After training completes, the machine is asked to cluster, predict or classify similar data. There are numerous techniques of machine learning, the most popular among those are ANN or Artificial Neural Network, SVM or Support Vectors Machine, and DT or Decision Trees, etc. Then there are combinations of these techniques which are called ensemble techniques.

2.2.1 Using Machine Learning to Benefit from Social Media Data

As stated in the previous section, data has been generated in abundance on Social Media and numerous researches have been carried out in order to benefit from the stated data. In this section, we discuss a few such studies which have made use of machine learning techniques and benefitted from the data produced on social media.

2.2.1.1 Face Recognition

In the past two decades, Face Recognition has become the crux of many new innovations and has been proffering steadily many cross-domain apps ranging from conventional software that is commercial to crucial applications for law enforcement. The newest innovative advancements in Analytics for Big Data, Machine Learning, Cloud Computing, and Social Networks have greatly changed the typical perception of how many daunting challenges in Computer Vision can be taken care of. In this research, an insight of the perceptions and ideas related to Cloud Computing, Social Networks, Big Data and Machine Learning from the viewpoint of FR has been provided. The stated study has also proffered a novel framework for FR founded on Extreme Learning Machines methodology in order to carry out the job of Face Tagging for Social Networks that operate on Big Data. In the approach that has been presented in this study, the stated concepts have been merged in order to develop a technique that is used to supplement FR's performance, moreover, it can also serve in other disciplines as well.

This study has been successful in proffering an insight view of the latest advancements in Social Networking, Big Data and Machine Learning and the different methods in which they can be converged with FR. The study has successfully proposed a new approach for FR which is more robust than all those conventional techniques which are not based on the cloud. It can garner from the previously presented techniques when it comes to boosting in accuracy and improvement in performance. The stated framework has been evaluated for Face Tagging task in Social Networks that operate on Big Data by making use of the Extreme Learning Machines procedure. As stated earlier this technique can also be used for other FR tasks for instance authentication and access control with more ease [12].

2.2.1.2 User Classification

This study has been carried out for classifying users in Social Media, and particularly on Twitter. The research is able to automatically deduce the values of attributes of users such as ethnicity or political orientation by leveraging information which can be observed for instance network structure, user behavior, and contents of the language of the feed of Twitter's users. A machine learning technique has been used which depends on a wide-spread range of features which are derived from the stated information of Twitter users. The technique has been trialed for 3 tasks having various properties and the results of the experiments are encouraging.

These tasks are the detection of affinity for a specific business, detection of political affiliation, and identification of ethnicity. The findings of this particular study reveal that rich linguistic features provide value across all the 3 tasks and can be termed as more promising as far as additional user identification is concerned [13].

2.3Deep Learning

Deep learning and machine learning are closely related and can be differentiated based on the fact that in deep learning we assume that we are not limited by machine resources. Since hardware resources are not limited, we do not opt for feature selection or feature reduction techniques but instead, feed all the features and attributes to our system and make use of them for reaching our desired results.

2.3.1 Using Deep Learning to Benefit from Social Media Data

Just like machine learning, deep learning has also been applied in order to make use of and to benefit from the abundantly produced and accessible social media data. In this section, we discuss a few studies which have applied deep learning techniques for the stated purpose.

2.3.1.1 Sentiment Analysis

Analyzing sentiments of contents that are generated by online users is significant for numerous social media analytics errands. Researchers mainly rely on sentiment analysis of texts in order to make systems for predicting elections (political) and for measuring economic indicators and many more such tasks. Lately, users of social networking sites have been using videos and images in order to express their viewpoints and share their experiences.

If these largely posted visual contents are analyzed for sentiments, it can help better when it comes to extracting sentiments of users regarding an event or a topic, for instance in those tweets (images and videos), so that predicting sentiments from these visual contents becomes complementary as far as analysis of sentiments from textual contents is concerned. This study is motivated by the requirements of leveraging hugely scaled training data that is noisy for solving highly difficult issues of sentiment analysis from images and makes use of CNN or more commonly known as Convolutional Neural Network. Firstly, an adequate architecture of CNN has been designed for analyzing sentiment in images. Over half a million training samples are acquired by making use of a baseline sentiment algorithm for labeling Flickr pictures. In order to utilize these noisy data that is labelled by machine, a progressive strategy is employed for fine-tuning the network. Moreover, the performance of the proposed technique is further improved on Twitter images by bringing a domain transformation by means of a tiny number of Twitter images that are manually labeled.

The experiments conducted on Twitter images having manually labeled images is quite extensive. The results of these experiments depict that the presented CNN can be used to achieve an ameliorated performance as far as analyzing sentiment in images is concerned [14].

2.3.1.2 Psychological Stress Detection

It is of utmost significance to discover or identify and manage stress before it starts causing severe issues. But it is also a fact that the currently present techniques of stress detection mostly depend on psychological devices or psychological scales, and hence not only making the process complicated but expensive as well. In this study, an effort of automatically detecting an individual’s psychological stress from social media has been put forth. The study makes use of real-time online micro-blog data, firstly a correlation between the stress of users and the content of their tweets, behavior patterns, and social engagement is investigated. Secondly, two types of attributes related to stress are defined i.e. low-level content attributes obtained from one tweet, consisting of social interactions, images and texts, and the second obtained from their weekly micro-blog, leveraging info such as time, and type of tweet, and the style of linguistics. Thirdly, CNN is designed for combining the content of attributes with attributes of stats. The CNN has cross auto-encoders used for generating user-scope content properties from content attributes that are low level. Lastly, a DNN or Deep Neural Network model has been proposed for incorporating the stated types of user-scope attributes for detecting the psychological stress of users.

The trained model is tested on four datasets obtained from different micro-blogging websites including Twitter and Tencent Weibo. The results obtained from different experiments reveal that the technique proposed is efficient and effective when it comes to identifying psychological stress from the stated micro-blogging data. The proposed model can be of use for the development of stress detection tools for health agents and other such individuals [15].

2.3.1.3 Traffic Flow Prediction

The timely and accurate traffic flow information is extremely vital for deploying systems of transportation that are intelligent. In the last decade or so, traffic data, just like other data, has been produced in abundance and the period of big data for transportation has truly upon us. The currently available methods of traffic flow prediction basically make use of traffic prediction models that are shallow and do not fulfill the requirements of numerous real-time applications. The stated scenario motivated researchers to rethink the prediction of traffic flow and to propose prediction models that are based on big traffic data. In this study, a new method for the prediction of traffic flow based on deep learning has been proposed, which cogitates the temporal and spatial correlations inherently.

A greedy layer-wise fashion-based training method is used with a stacked autoencoder model for learning the features of a generic flow of traffic. The authors claim that this the first time a deep architecture has been deployed with autoencoders as building blocks for the representation of traffic flow features. The researchers also claim that their proposed technique has outclassed the previous techniques as far predicting traffic flow is concerned [16].

2.4Natural Language Processing

Natural Language Processing or more commonly known as NLP can be termed as a subfield of information or data engineering, computer sciences, and artificial intelligence. NLP mainly deals with the communications or interactions carried out between human and computer languages, particularly with the programming of computers for processing and analyzing humongous amounts of data that is present in natural languages i.e. human languages.

2.4.1 Using Natural Language Processing to Benefit from Social Media Data

Just like Machine learning and deep learning techniques, NLP is also widely used for making use of data that is currently being produced on Social Media platforms such as Twitter and Facebook, etc. In this section, we look at a few studies which have done the stated job.

2.4.1.1 Enhancing Emergency Situation Awareness

Platforms of Social Media, for instance, Twitter, provide a great source of real-time information regarding events that happen in real-world, specifically when mass emergencies take place. The examination and selection of cherished information obtained from Social Media give a detailed vision regarding time crucial scenarios to emergency personnel to better comprehend the influence of dangers and perform emergency procedures timely. This particular research brings forth an analysis performed on Twitter messages that are generated amid natural catastrophes and depicts how the technique of NLP and data mining can be made use of for extracting situation mindfulness details from Twitter. The stated research delivers fundamental methods that have been examined these methods include tweet filtering and identification, burst detection, geotagging and online clustering.

The rapidly evolving and increasing usage of Social Media amid natural crises and disasters gives us the opportunity to gain information reported on the ground from the public. This study emphasizes the analysis of Twitter messages that are generated amid natural disasters and humanitarian crises and brings forth fundamental approaches for detecting bursts, filtering, and classification of tweets, geotagging and clustering. The evaluation and development of the stated approaches and methods revealed that given that the right information is chosen via Social Media, it can facilitate the concerned authorities and personnel for enhancing their attentiveness of time crucial scenarios and taking better decisions as responses to emergencies [17].

2.4.1.2 Detecting Poor Quality Healthcare

Just like other service providers, the healthcare sector has also been seeing an increase in interest of patient-centered care and calls in order to improve the patient’s experience. Simultaneously, numerous patients have been making use of the internet for describing their involvements with healthcare. The authors of this paper believe that the increase of access to patients' blogs, social networking sites like Twitter or any other review systems or sites provides the concerned person with the opportunity of advancing the agenda of making their service patient-centric. Moreover, a good amount of data is also generated at the same time. The stated concept has been described by the authors as 'a cloud of patient experience'. In the paper under discussion, the methods in which the combination of patients' explanations of their own experience on the Internet can be utilized for detecting poor clinical care. With the passage of time, the stated technique can also be used to recognize excellence and empower to be built on.

The study suggests that methods of NLP and sentiment analysis should be used to change the descriptions of patients' experiences that are not structured into useful events of healthcare performance [18].

2.5Twitter Gender Classification

If the gender of Twitter users is classified correctly, the information can be used for numerous purposes i.e. for directing particular advertisements to one or the other gender, the stated information can also be used by law enforcement for legal investigation and also by others for social reasons and benefits.

2.6A Survey of Machine Learning and Deep Learning Techniques Used for Twitter Gender Classification

One of the studies from the literature aims to classify twitter gender. The technique proposed in the stated study is based on Support Vector Machine (SVM). It has used CLEF corpus for training and testing and has been able to achieve 81.7%. A deep analysis and insight of the article reveal that if deep neural networks would enhance the classification accuracy of the proposed system [19].

Another study from the literature tries to profile users' demography by making use of a technique founded on logistic regression it uses CLEF and achieves 69.17%. It is evident that the achieved accuracy can be enhanced to a great deal, and hence further experimentations ought to be performed [20].

A study in the literature aims to reduce the number of features for empirically evaluating the characteristics of twitter profiles for classifying gender. This study has used many techniques and has discovered that Decision Tree works best for their dataset of around 194,293 twitter profiles. The maximum accuracy attained by this study is close to 70%, thus further work should be carried out in order to improve the stated accuracy [21]. An interesting study tries to classify gender of twitter profiles and achieves a decent accuracy of 89.5% and contributes to the literature by generating new features. The stated article makes use of Twitter API in order to cumulate a new corpus [22].

Gender classification of users based on a corpus cumulated from twitter API has been carried out in another study. This has made use of numerous classification techniques and recognizes Best First Tree as the best technique, which has achieved an accuracy of 81.66% [23].

Author’s gender classification of a dataset assembled from Twitter Streaming API has been made use of and a stream-based neural network has been used to achieve an accuracy of around 98% [24]. [26] also uses the Twitter Streaming API to get twitter data and achieves similar accuracy as [24].

Recognition of gender from Twitter profiles has been conducted in another study, and an ensemble technique based on the convolutional neural network has been proposed in this study. A corpus of 274,933 New Yorkers has been used in this study. The technique has been able to achieve an accuracy of 86.54%. For future new traits or attributes can be taken into consideration in order to enhance the performance of the ensemble [25].

Adience Benchmark has been used for classifying the age and gender of twitter users by a study. It has proposed a technique founded on CNN and has achieved an accuracy of 87% [27]. Classification of gender from tweets by a technique proposed in an article based on Balanced Winnow. The proposed technique has been able to achieve an accuracy of 92% [28].
One study from the literature aims to automatically detect twitter users’ gender by trying different techniques and Fuzzy C-means is recognized as the best technique [29].

Language independent Twitter users’ gender classification has been carried out in another study, which has proposed a technique based on SVM. The study has used TwiSty as its corpus. Users of this dataset have self-reported their gender and hence the proposed technique has to be evaluated on another dataset [30].

TIRA dataset has been used to recognize the age, gender, and personality of users. The study under scrutiny has used a linear model with stochastic gradient descent. The study has been successful 60% of the time [31].

Distributed K-means clustering has been used by a study for twitter user classification on 2.2 billion tweets to achieve an accuracy of over 90% [32]. Gender is inferred from tweets of 5,000 users by a study. The stated study has proposed a technique based on SVM and has achieved 83% accuracy [33]. Another study from the literature aims to identify the gender of Twitter users, the technique presented in the study is based on SVM Light. The study has been able to achieve 80% [34].

Twitter users' demographics have been predicted by a study that makes use of regression models. The corpus used in this study has been derived from Quandcast and it has achieved a mediocre accuracy of 69% [35]. Twitter users' type of recognition is the aim of another study in the literature. The study founds its technique on SVM with a linear kernel. The corpus has been assembled through twitter streaming API and the technique is right 89% of the times [36].
Semantic analysis is used by a study to predict twitter users' gender with the help of SVM. 10,000 profiles are considered as the corpus, and 88% accuracy has been achieved by the study. In the future, the authors intend to make use of visual information for predicting age and political association [37].

Twitter users' classification has been done by a study which uses CNN with multiple classifiers i.e. Decision Tree, SVM, AdaBoost, Gradient Boosting, & Random Forest. CrowdFlower dataset has been used for training and testing, and the study has achieved 89.9% accuracy. In the future, the researchers want to use more features and deploy deep neural networks for enhancing feature fusion [38].

User types are inferred as males, females and organizations by a study. CNN has been utilized for image learning whereas for classification different techniques are used in this study [39]. Age and gender recognition are the aims of research, and numerous classification techniques have been tried, and Extra Trees (1,000 classifiers with stop words) has outperformed all other techniques.

The study has used CrowdFlower dataset and has achieved 91.4% accuracy [40]. A study from the literature uses CNN to predict twitter gender. It uses PAN 2018 Author Profiling dataset and achieves 79% accuracy [41].

The following table i.e. table 2-1 is used to present a list of machine learning and deep learning techniques applied for twitter gender classification.

Table 2- 1: ML & DL techniques for Twitter Gender Classification

Sources	Research Problem	Proposed Approach	Corpus	Mean Accuracy in %	Future Work/Limitations
[19]	Twitter gender identification.	The proposed technique is based on Support Vector Machine (SVM).	CLEF	81.7	Application of deep neural networks may enhance the classification accuracy.
[20]	Twitter users' demographics and social profiling.	The presented technique is founded on logistic regression.	CLEF	69.17	The accuracy can be improved further.
[21]	This study aims to reduce the number of features for empirically evaluating the characteristics of twitter profiles for classifying gender.	They have used Decision Tree and many other techniques for classification purpose.	A dataset of 194,293 twitter profiles has been used in this study.	69.13	In the future, addition tweet and profile characteristics could be applied to enhance the technique.

[22]	Classification of gender in twitter profiles.	The main contribution of this study, is feature generation.	Twitter API is used to cumulate a corpus.	89.5	Limited dataset.
[23]	Gender classification of users based on Twitter data.	This research has used 3 classification techniques and recognizes Best First Tree as the best technique.	Twitter API is used to cumulate a corpus of 3,240 tweets.	81.66	Limited dataset
[24]	Author's gender identification.	The proposed technique is based on a stream-based neural network.	Twitter Streaming API has been used and to assemble a corpus of over 36,000.	97.98	-
[25]	Recognizing gender from Twitter profiles.	An ensemble technique based on a convolutional neural network has been proposed in this study.	Corpus of 274,933 New Yorkers has been used in this study.	86.54	New traits or attributes can be taken into consideration in order to enhance the performance of the ensemble.

[27]	Classifying the age and gender of Twitter users.	The proposed technique has been derived from CNN.	Adience Benchmark.	86.8	More training data can be used to improve the performance of the algorithm.
[28]	Discriminating gender from tweets.	The classifier is derived from Balanced Winnow.	Twitter's API has been used to assemble a corpus of 213 million tweets.	92	-
[29]	Automatic detection of Twitter users' gender.	Different techniques have been trialed in this study and Fuzzy C-means is recognized as the best technique.	Tweets of 242,000 Twitter users.	96	Development of a semi-automatic system for labelled dataset.
[30]	Language independent Twitter users' gender classification.	The proposed technique uses SVM as a classifier.	TwiSty Dataset.	-	Users of this dataset have self-reported their gender and hence the proposed technique has to be evaluated on another dataset.
[31]	Recognition of age, gender and personality.	The proposed approach is based on a linear model	TIRA	60	The accuracy can improved be further.

		with stochastic gradient descent.			
[32]	Twitter user classification	The presented technique is based on distributed K-means clustering.	2.2 billion Tweets.	90.2	-
[33]	Inferring gender from Twitter data.	The proposed approach is based on SVM.	Tweets of 5,000 users.	83	Adding more features to enhance the accuracy.
[34]	Twitter profile gender identification.	This study uses SVM Light for classification.	Twitter search API, Nigerian data.	80	The technique can be applied to data that is not labelled.
[35]	Predicting Twitter users' demographics.	This study makes use of regression models for classification.	Data derived from Quandcast	69	In future, the generalization of the proposed technique can be evaluated.
[36]	Twitter users' type recognition	SVM with a linear kernel has been deployed in this study.	Corpus assembled through Twitter Streaming API.	89	-
[37]	Prediction of Twitter users' gender on the basis of semantic analysis.	SVM with RBF kernel has been utilized in this research.	10,000 Twitter profiles.	88	The authors intend to make use of visual information for predicting age

					and political association.
[38]	Role-based Twitter user's classification.	CNN with multiple classifiers i.e. Decision Tree, SVM, AdaBoost, Gradient Boosting, & Random Forest.	CrowdFlower & Gender Labelled Dataset	89.9	The researchers want to use more features and deploy deep neural networks for enhancing feature fusion.
[39]	Inferring user types as males, females and organizations.	CNN has been utilized for image learning whereas for classification different techniques are used in this study. For CrowdFlower Random Forest has achieved the highest accuracy & for ILLAE SVM does best.	CrowdFlower & ILLAE	85.99 & 78.61	Profile pictures of organizations could not be recognized and female with short hair are identified as male.
[40]	Gender and Age recognition of Twitter users.	Numerous classification techniques have been tried, and Extra Trees (1,000 classifiers with stop words) has outperformed	CrowdFlower	91.4	For gender classification an ensemble is intended to be applied.

		all other techniques.			
[41]	Twitter gender prediction	The proposed technique has made use of convolutional neural networks.	PAN 2018 Author Profiling dataset	79	The researchers want to make use of images that are present in the dataset.

This chapter of the manuscript starts with the definition of Social Media, then it proceeds toward the amount of data produced on Social Media platforms such as Twitter, Facebook, Instagram, and Flickr, etc. Then the different techniques of machine learning, deep learning and natural language processing that are applied to the abundant data produced on the stated platforms in order to benefit from them are discussed. Then a survey regarding the main theme i.e. machine learning and deep learning techniques applied for Twitter gender identification or classification has been presented in the shape of a table. The main findings of numerous studies have been brought forth in this table. The future work/limitation column can be used for advancing a particular study of the literature.

CHAPTER 3

PROPOSED METHODOLOGY

In this chapter, the methodology of the thesis has been discussed. The discussion begins with an illustration of the dataset that has been used for both validation and testing purposes. Then preprocessing has been described briefly in which emojis, and special characters, etc. are dealt with. Due to the fact that character embeddings take a shorter time than word embeddings, and that is the reason why character embedding has been preferred over word embedding. The chapter concludes by providing an insight into the architecture of CNN and by providing a figurative illustration of the proposed methodology.

3.1 Data

The dataset that has been used in this thesis is known as the CrowdFlower. The dataset contains 20,000 rows, each with a username, a random tweet, account profile and image, location, and even link and sidebar color. Because of multiple annotation results by different people, each row has a certain degree of confidence signifying the likelihood of the majority gender. It has been made sure to use only the cleanest data possible, user types have been filtered every user type that had a degree of confidence less than 1. After filtering the dataset, there are 13,926 users. Because most of the user's image URL is not functional. In the end, 10,029 users could be retrieved and used as training samples. The distribution of the CrowdFlower dataset is presented in the following table i.e. table 3-1 [28].

Table 3- 1: Dataset with User Distribution

Dataset	Male	Female	Total
CrowdFlower	4,658	5,371	10,029

3.2 Preprocessing

In Twitter, characters are used not only to create words but also to express emotions like smiling as ':)' or blinking as ';)', because of this type of usage, punctuations and stop words did not get eliminated, texts are given as how they are. NLTK [10] is used to tokenize tweets. To illustrate (example from NLTK):

Tweet = "This is a cooool #dummysmiley: :-) :-P <3 and some arrows < > -> <–"

Tokenized Tweet = [’This’, ’is’, ’a’, ’cooool’, ’#dummysmiley’, ’:’, ’:-)’, ’:-P’, ’<3’, ’and’, ’some’, ’arrows’, ’<’, ’>’, ’->’, ’<–’]

Each word in the tokenized tweet is applied lowercasing. Then, each character from the word is taken to be utilized in the input to the system. Thus, the tweet in the above example is turned into the following input:

Input = [’t’, ’h’, ’i’, ’s’, ’i’, ’s’, ’a’, ’c’, ’o’, ’o’, ’o’, ’o’, ’l’, ’#’, ’d’, ’u’, ’m’, ’m’, ’y’, ’s’, ’m’, ’i’, ’l’, ’e’, ’y’, ’:’, ’:’, ’-’, ’)’, ’:’, ’-’, ’p’, ’<’, ’3’, ’a’, ’n’, ’d’, ’s’, ’o’, ’m’, e’, ’a’, ’r’, ’r’, ’o’, ’w’, ’s’, ’<’, ’>’, ’-’, ’>’, ’<’, ’-’, ’-’]

For each user, the number of characters is set to the highest number that is allowed for tweets in Twitter. If a tweet has a fewer number of characters than the maximum, padding is applied to the end of the tweet.

3.3Character Embeddings

Character embeddings with size 25 are initialized by sampling from a uniform distribution with 0 mean and trained simultaneously with the neural network. Due to their smaller size and count, training character embeddings requires fewer text to be trained than word embeddings. Therefore, the given dataset was sufficient to train them and no additional data are collected or used.

3.4Architecture

In the proposed technique each tweet of a user is passed to the CNN simultaneously as a sequence of characters to assess the style-based features of each particular tweet. CNN outputs a feature vector for each tweet.

At this level, using other methods like combining, flattening or averaging the feature vectors would mean to explicitly assume the equal importance among tweets. However, the level of information on gender may differ from tweet to tweet. Therefore, A Bahdanau attention mechanism [41] is combined with the character CNN in order to learn which tweet holds more information on the gender of its author. Figure 3-1 shows the attention mechanism in detail which is calculated by the following equations:

$$\mathrm{Ai} = \tanh(\mathrm{W} \propto \mathrm{ti} + \mathrm{b})$$

$$\mathrm{vi} = \frac{\exp(\mathrm{Ai}\omega\mathrm{i})}{\sum jexp(Aj\omega j)}$$

$$Oi - viti$$

$$K = \sum_{i} Oi$$

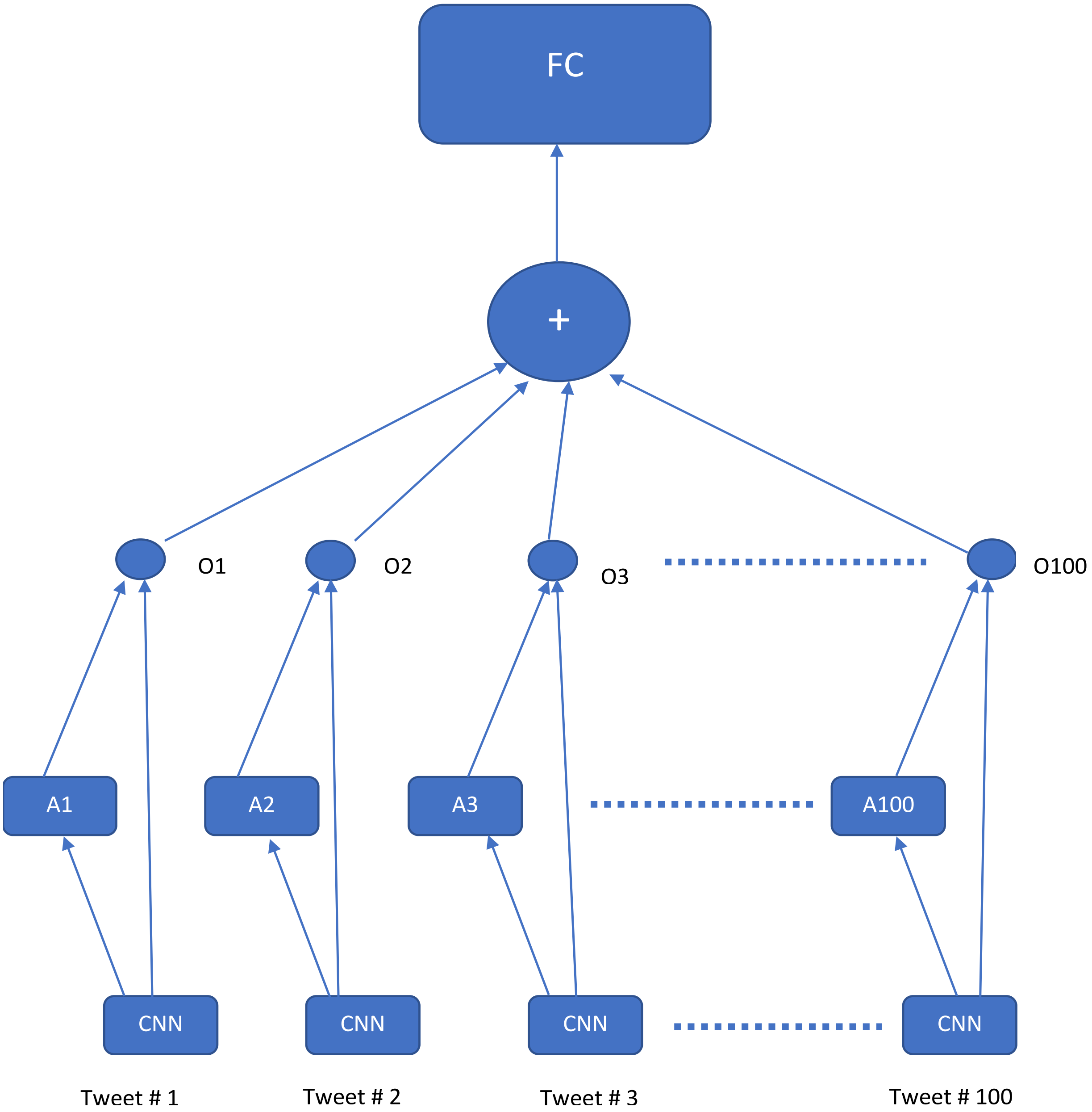


Figure 3- 1: Attention Mechanism

A fully connected layer is used on the output of the attention layer to reduce the size of the feature vector to the number of genders. Predictions are obtained after applying Softmax over the output of the fully connected layer. The proposed model can be seen in Figure 3-2.

CNN [6] is implemented with ReLu activation function and [filter size, embedding size] shaped filters with stride 1 to make all characters visited. Adam optimizer [42] is used with cross-entropy loss. To prevent the model from overfitting L2 regularization loss is used.

3.1Parameter Selection

Exhaustive grid search is used to optimize the hyperparameters of the model. Parameters that have been tried for each language can be seen in Table 2. Due to differences in each language and the size of the dataset, different hyperparameters gave best results for each language (Table 3-2).

Table 3- 2: Hyperparameter used in Optimization

Parameter	**Values**
Embedding Size	25
Learning Rate	5x10^3, 10^4, 5x10^4, 10^5, 5x10^5, 10^6
L2 Regularization Coefficient	5x10^4, 10^5, 5x10^5, 10^6, 5x10^6, 10^7, 5x10^7, 10^8
Filter sizes	3, 4, 5
Number of Filters	40, 50, 60, 75, 100

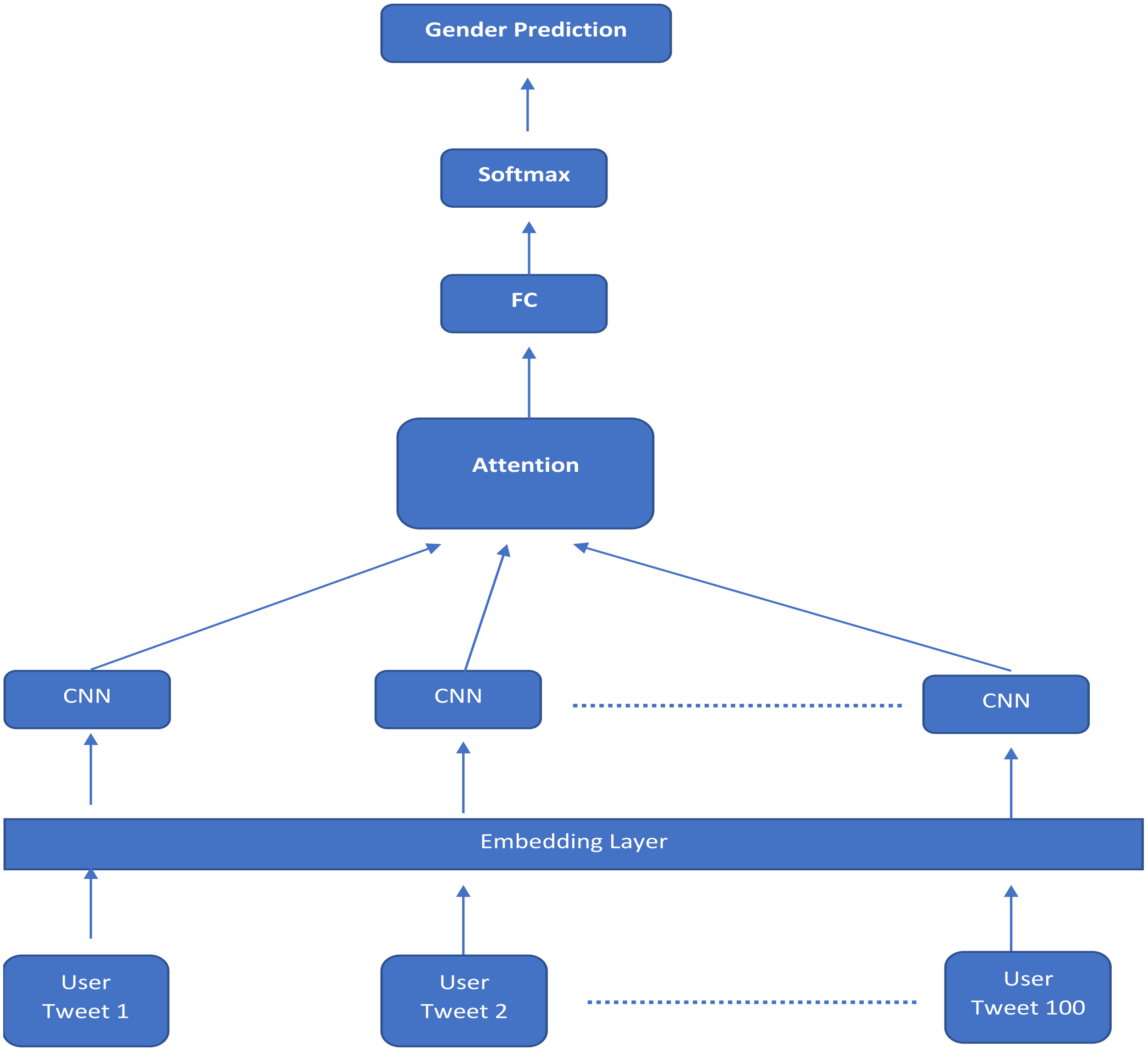


Figure 3- 2: Proposed Methodology

CHAPTER 4

IMPLEMENTATION

This chapter of the thesis discusses the implementation of the proposed technique and hence can be termed as the most significant chapter of the manuscript.

4.1Convolutional Neural Network

When we hear about Convolutional Neural Network (CNNs), we typically think of Computer Vision. CNNs were responsible for major breakthroughs in Image Classification and are the core of most Computer Vision systems today, from Facebook's automated photo tagging to self-driving cars. More recently we have also started to apply CNNs to problems in Natural Language Processing and gotten some interesting results. In order to understand convolution, the easiest way is by thinking of it as a sliding window function applied to a matrix.

CNNs are basically just several layers of convolutions with *nonlinear activation functions* like ReLU or Tanh applied to the results. In a traditional feedforward neural network, we connect each input neuron to each output neuron in the next layer. That is also called a fully connected layer, or affine layer. Whereas in CNNs that is not done, instead, convolutions are used over the input layer to compute the output. This results in local connections, where each region of the input is connected to a neuron in the output. Each layer applies different filters, typically hundreds or thousands like the ones showed above, and combines their results. There's also something called pooling (subsampling) layers. A key aspect of Convolutional Neural Networks is *pooling layers,* typically applied after the convolutional layers. Pooling layers subsample their input. The most common way to do pooling is to apply a max operation to the result of each filter. During the training phase, CNN automatically learns the values of its filters based on the task you want to perform. For example, in Image Classification, a CNN may learn to detect edges from raw pixels in the first layer, then use the edges to detect simple shapes in the second layer, and then use these shapes to deter higher-level features, such as facial shapes in higher layers. The last layer is then a classifier that uses these high-level features.

There are two aspects of this computation worth paying attention to: Location Invariance and Compositionality. Let’s say you want to classify whether or not there’s an elephant in an image. Because you are sliding your filters over the whole image you don’t really care *where* the elephant occurs. In practice, *pooling* also gives you invariance to translation, rotation and scaling, but more on that later. The second key aspect is (local) compositionality. Each filter *composes* a local patch of lower-level features into higher-level representation. That’s why CNNs are so powerful in Computer Vision. It makes intuitive sense that you build edges from pixels, shapes from edges, and more complex objects from shapes.

As far as Natural Language Processing is concerned, instead of image pixels, the input to most NLP tasks are sentences or documents represented as a matrix. Each row of the matrix corresponds to one token, typically a word, but it could be a character. That is, each row is vector that represents a word. Typically, these vectors are *word embeddings* (low-dimensional representations) like word2vec or GloVe, but they could also be one-hot vectors that index the word into a vocabulary. For a 10word sentence using a 100-dimensional embedding, we would have a 10×100 matrix as our input. That’s our “image”.

4.2CNN Architecture

An illustration of the architecture of CNN used in this manuscript has been provided in figure 4-1.

4.2.1 Initializing CNN

In this stage, the convolutional neural network is initiated.

4.2.2 Embedding layer

Embedding layer is used for neural network on text data and defined as a first hidden layer of the network. It specifies three arguments input dimension, output dimension, and input length.
The reasons to have an embedding layer are:

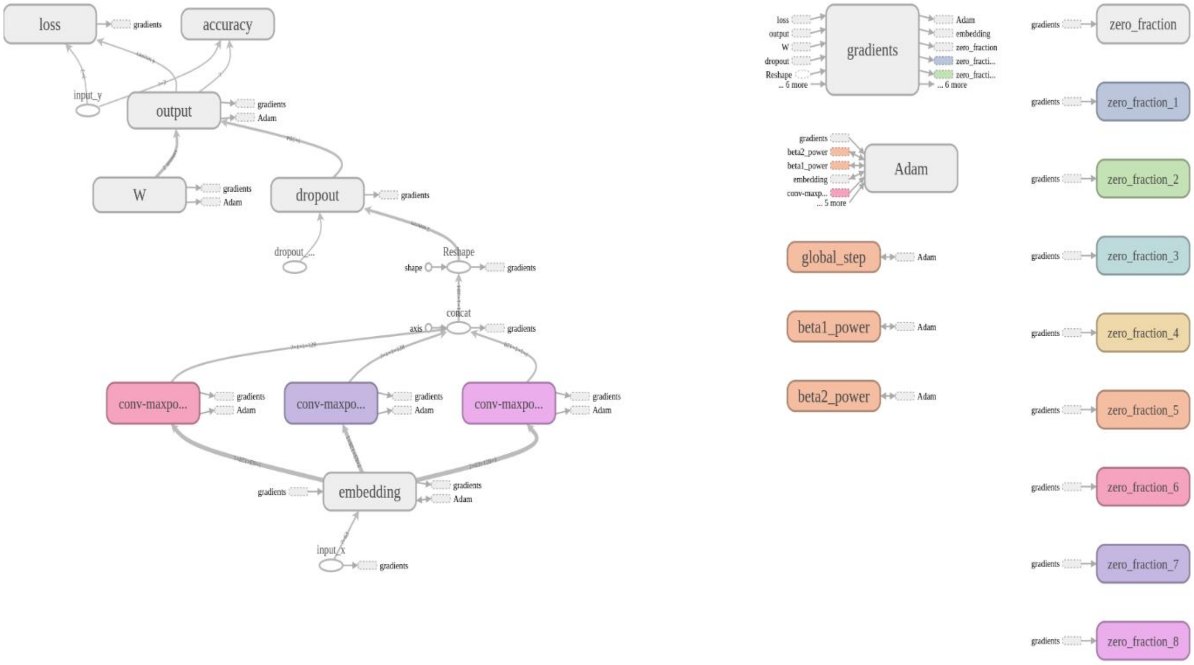


Figure 4- 1: CNN Architecture

1. One-hot encoded vectors are high-dimensional and sparse. Let's assume that we are doing Natural Language Processing (NLP) and have a dictionary of 2000 words. This means that, when using one-hot encoding, each word will be represented by a vector containing 2000 integers. And 1999 of these integers are zeros. In a big dataset, this approach is not computationally efficient.

2. The vectors of each embedding get updated while training the neural network. If you have seen the image at the top of this post you can see how similarities between words can be found in a multi-dimensional space. This allows us to visualize relationships between words, but also between everything that can be turned into a vector through an embedding layer [43].

The architecture that has been chosen after experimentations and which is the best suited for different rounds of validations has three 4D convolution layers.

Adding the first convolution

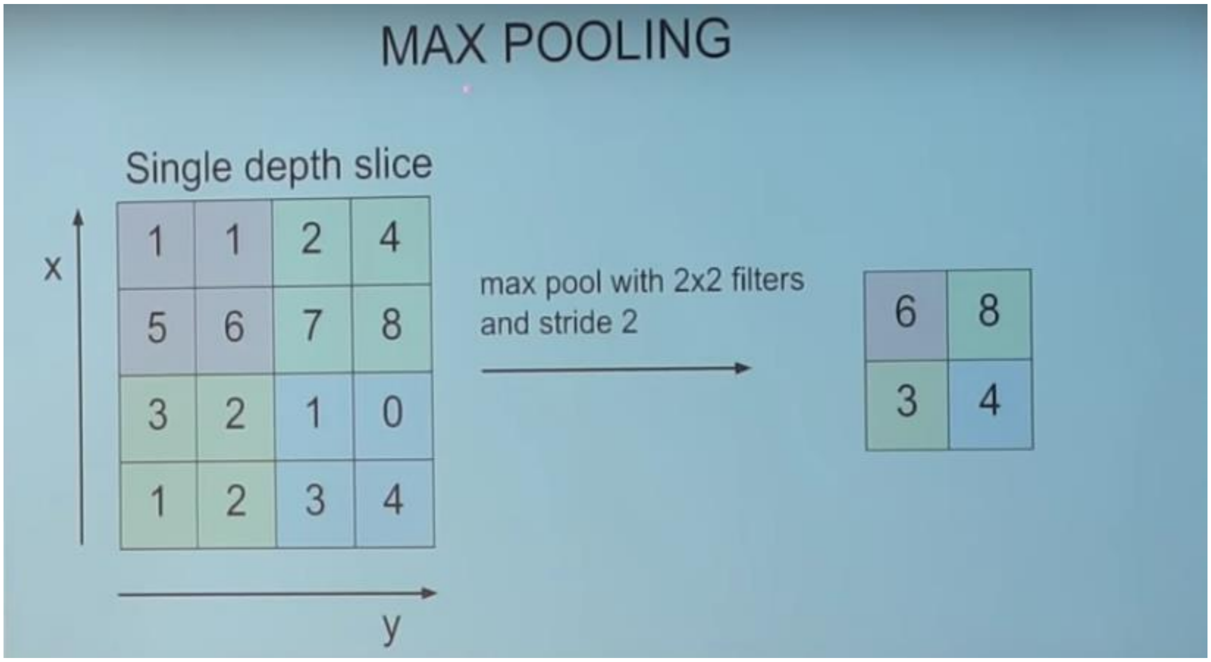
Step 1 convolution

A convolution is how the input is modified by a filter. In convolutional networks, multiple filters are taken to slice through the image and map them one by one and learn different portions of an input image. Imagine a small filter sliding left to right across the image from top to bottom and that moving filter is looking for, say, a dark edge. Each time a match is found, it is mapped out onto an output image [44].

- Input shape= (3, 128, 1, 128)
- Activation = ReLu

Step 2 Maxpooling

A pooling layer is another building block of a CNN. Its function is to progressively reduce the spatial size of the representation to reduce the number of parameters and computation in the network. Pooling layer operates on each feature map independently. The most common approach used in pooling is max pooling. Figure 4-2 provides a graphical representation of how Maxpooling is carried out [45].



MAX POOLING

Single depth slice

1	1	2	4
5	6	7	8
3	2	1	0
1	2	3	4

max pool with 2x2 filters and stride 2

6	8
3	4

Figure 4- 2: Maxpooling

Adding the second convolution

Step 1 convolution

- Input shape= (4, 128, 1, 128)
- Activation = ReLu

Step 2 Maxpooling

Adding the third convolution

Step 1 convolution

- Input shape= (5, 128, 1, 128)
- Activation = ReLu

Step 2 Maxpooling

4.2.3 Adding Concat layer

This layer is used to concatenate a list of inputs. It takes as input a list of tensors, all of the same shape except for the concatenation axis, and returns a single tensor, the concatenation of all inputs. It receives the following arguments.

- **axis**: Axis along which to concatenate.
- ****kwargs**: Standard layer keyword arguments [46].

4.2.4 Adding reshape layer

In Keras, this is a typical process for building a CNN architecture: Reshape the input data into a format suitable for the convolutional layers, using X_train.reshape() and X_test.reshape() [47].

4.2.5 Adding dropout layer

Dropout is a technique where randomly selected neurons are ignored during training. They are “dropped-out” randomly. This means that their contribution to the activation of downstream neurons is temporally removed on the forward pass and any weight updates are not applied to the neuron on the backward pass.

As a neural network learns, neuron weights settle into their context within the network. Weights of neurons are tuned for specific features providing some specialization. Neighboring neurons become to rely on this specialization, which if taken too far can result in a fragile model too specialized to the training data. This reliant on the context for a neuron during training is referred to as complex co-adaptations [48].

4.2.6 Compiling CNN

- Optimizer = Adam

Usually, forward propagation and backward propagation techniques are used as error functions or weight optimization. But in our case, we use the Adaptive moment estimation technique. However, Adam has been used as an optimization technique in this system.

Since 2014, a special optimization algorithm in the shape of Adam (Adaptive Moment Estimation) for deep neural networks is present [7]. Adam is one of the best methods that are used to calculate adaptive learning rates for every parameter. It computes the adaptive learning rates for all the parameters. Apart from storing the exponentially decaying averages of previously squared gradients, for instance, RMSprop and Adadelta, it also keeps something similar to momentum. If momentum is thought of like a ball going down a slope, Adam can be termed as a heavy ball having friction and hence providing us with flat minima [49].

- Loss= softmax_cross_entropy_with_logits_sg

The Softmax regression is a form of logistic regression that normalizes an input value into a vector of values that follows a probability distribution whose total sums up to 1. The output values are between the range [0,1] which is nice because we are able to avoid binary classification and accommodate as many classes or dimensions in our neural network model. This is why SoftMax is sometimes referred to as a multinomial logistic regression.

As an aside, another name for Softmax Regression is Maximum Entropy (MaxEnt) Classifier.

The function is usually used to compute losses that can be expected when training a data set. Known use-cases of SoftMax regression are in discriminative models such as Cross-Entropy and Noise Contrastive Estimation. These are only two among various techniques that attempt to optimize the current training set to increase the likelihood of predicting the correct word or sentence [50].

Softmax function can be illustrated as [51]:

$$f_j(z) = \frac{e^{z_j}}{\sum_k e^{z_k}}$$

Hand in hand with the SoftMax function is the cross-entropy function. Here's the formula for it:

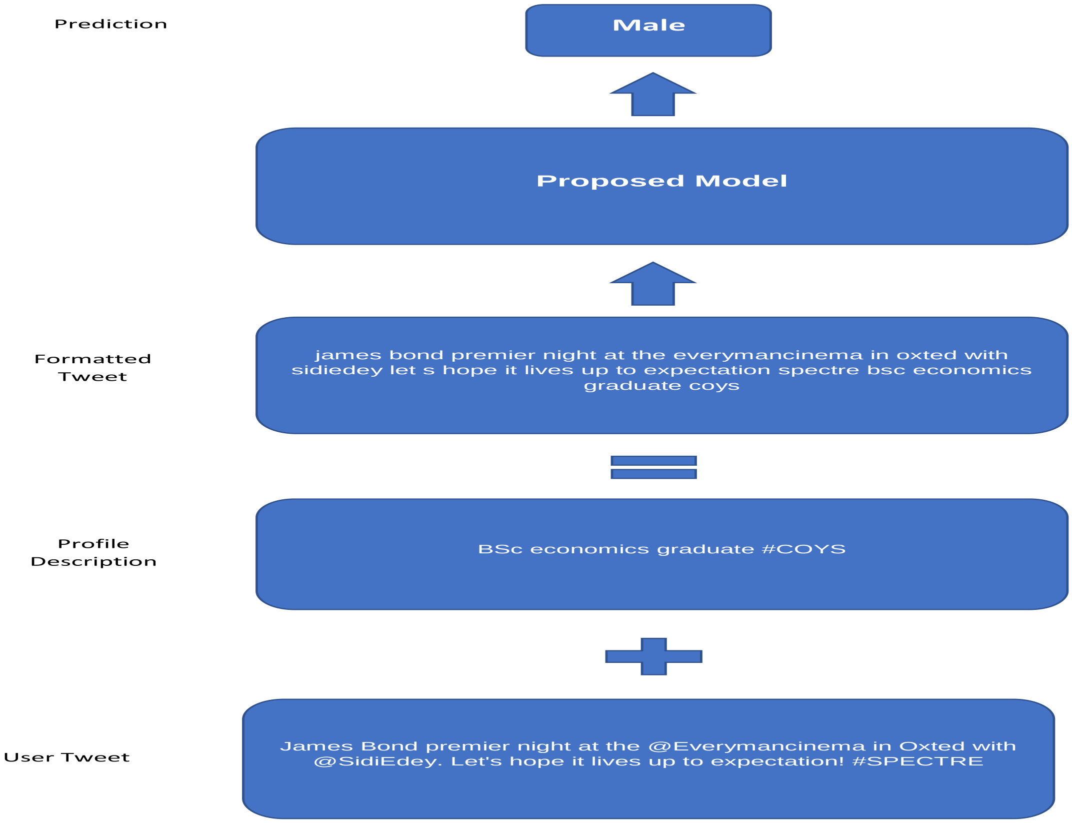
$$L_i = -\log\left(\frac{e^{f_{y_i}}}{\sum_j e^{f_j}}\right)$$

$$H(p,q) = -\sum_x p(x)\,\log q(x)$$

4.2.7 Prediction

This is the last stage of the process, and as it is clear from the heading in this stage the final prediction is made, which obviously is of whether the user is male or female.

The following figures i.e. figure 4-3 and figure 4-4 can be considered as sample tweets and their predicted genders.



james bond premier night at the everymancinema in oxted with sidiedey let s hope it lives up to expectation spectre bsc economics graduate coys

BSc economics graduate #COYS

James Bond premier night at the @Everymancinema in Oxted with @SidiEdey. Let's hope it lives up to expectation! #SPECTRE

Figure 4- 3: Sample Tweet and Prediction 1

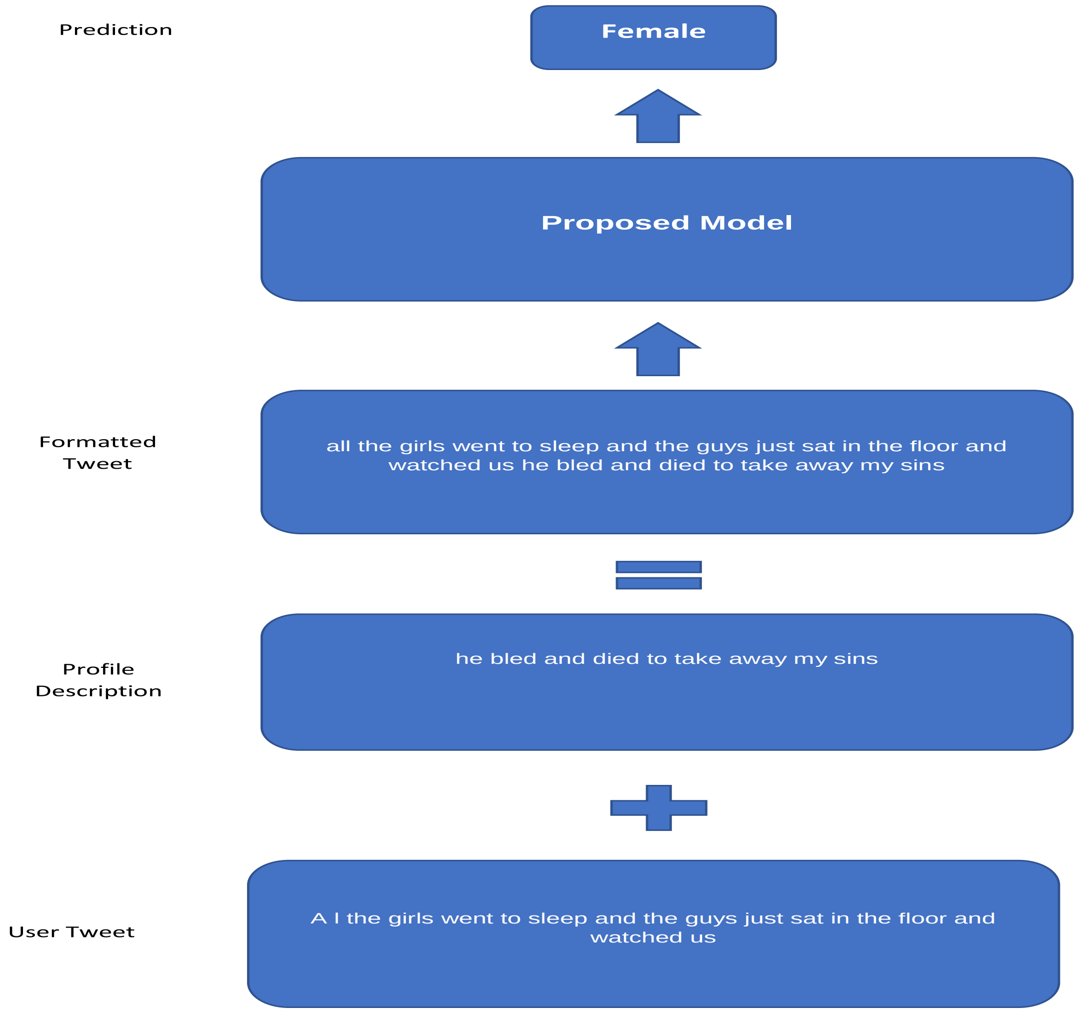

Prediction
Female
Proposed Model
Formatted Tweet
all the girls went to sleep and the guys just sat in the floor and watched us he bled and died to take away my sins
Profile Description
he bled and died to take away my sins
User Tweet
A l the girls went to sleep and the guys just sat in the floor and watched us

Figure 4- 4: Sample Tweet and Prediction 2

4.3Hardware

The system on which the technique has been executed has the following specifications.

- Processor: Intel® Core™ i7-7500U CPU @ 2.70GHz × 4
- Graphics: Intel® HD Graphics 620 (Kaby Lake GT2)
- OS Type: 64-bit
- RAM: 15.6 GiB

4.4Tools Used

As far as the tools and software that have been made use of are concerned, here is the list.

- Language: Python 3.5.2
- Anaconda

    - Libraries: Keras>TensorFlow (at the Backend)
- IDE: Spyder 3.2.3

In this chapter, the implementation of the proposed technique has been discussed in detail. An illustration not only graphic but textual has been presented in the same chapter. The architecture of CNN has been thoroughly explained. The illustration begins with the initiation of the network and continues until the prediction is done.

CHAPTER 5

RESULTS, COMPARISON, AND ANALYSIS

In this chapter, the reader is presented with an illustration of the results that have been achieved through the proposed technique. Moreover, a comparison of existing techniques applied on the same dataset and their accuracies has also been presented. As a subsection, an analysis of the proposed technique has also been presented.

5.1Results

The proposed technique has been able to achieve an accuracy of 97 percent. It is evident from the following figure i.e. figure 5-1 that the accuracy of the model has been increasing as the number of epochs increase.

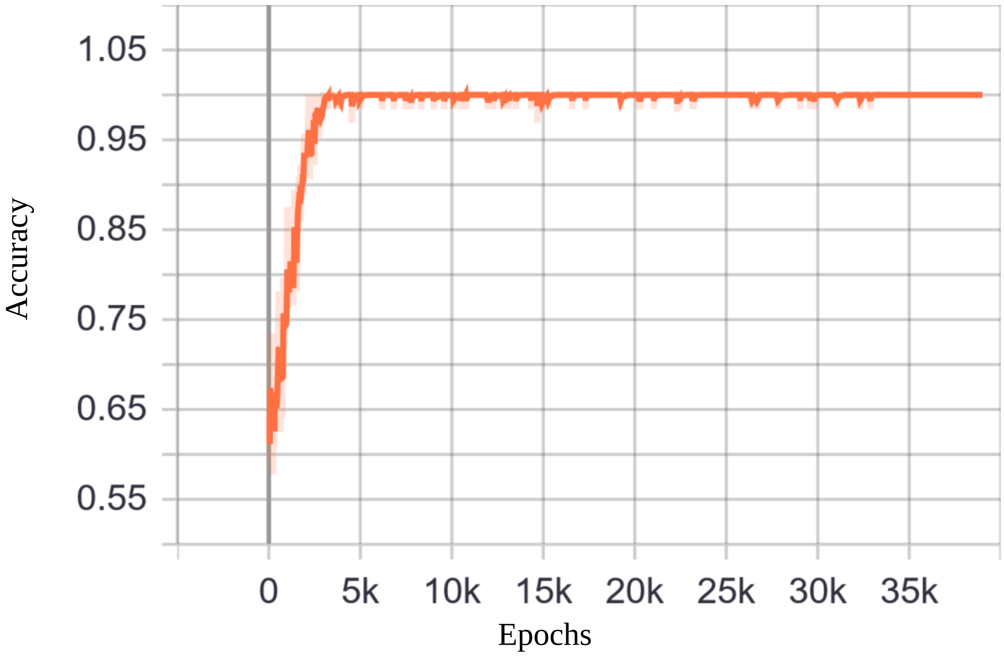


Figure 5- 1: Accuracy vs Epochs Graph

As far as loss is concerned, the objective is to minimize the loss function. The following figure i.e. figure 5-2 is used to depict the relationship between loss and epochs in the form of a graph.

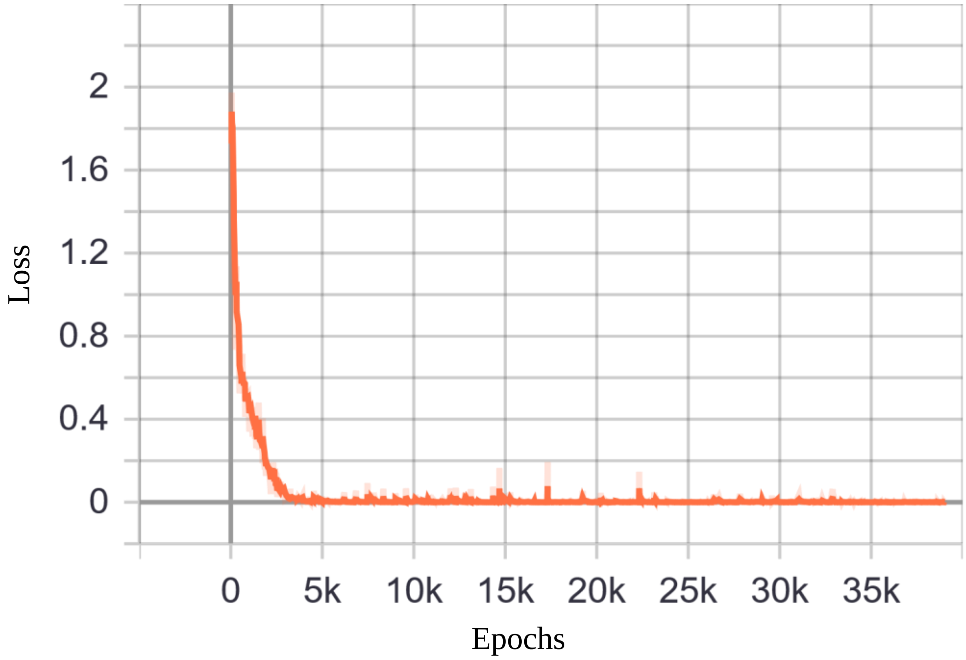


Figure 5- 2: Loss vs Epochs

In the following table i.e. table 5-1, the confusion matrix of the training process has been provided. Moreover, in order to get an insight of the effectiveness of the proposed model the sensitivity and specificity of the model have also been presented.

Table 5- 1: Confusion Matrix (Training)

	Predicted Positive	**Predicted Negative**
True Positive	8,918	241
True Negative	237	4,421

Sensitivity = True Positive/ True Positive+ Predicted Negative

= 8,918/8,918+237 = 0.97

Which means the sensitivity of the proposed model as far as the training process is concerned is 97%.

Specificity = True Negative/ True Negative+ Predicted Positive

= 4,421/4,421+241 = 0.94

The specificity of the proposed model as far as the training process is concerned is 94%.

Moreover, in the following table the confusion matrix of the testing process has been presented in the following table i.e. table 5-2.

Table 5- 2: Confusion Matrix (Testing)

	Predicted Positive	**Predicted Negative**
True Positive	490	9
True Negative	23	481

Sensitivity = True Positive/ True Positive+ Predicted Negative

= 490/490+23 = 0.95

Which means the sensitivity of the proposed model is equal to 95%.

Specificity = True Negative/ True Negative+ Predicted Positive

= 481/481+09 = 0.98

Which implies that the specificity of the proposed model is equal to 98%.

5.2State of the Art

In this section, a table (table 5-2) of studies that made use of the same dataset has been presented. The objective here is to compare the results i.e. accuracy of the proposed technique with those of existing state of the art techniques, in order to fetch an insight regarding the performance of the presented system.

Table 5- 3: Comparison with State of the Art

Source	Problem Statement	Proposed Technique	Dataset	Accuracy	Future Work/ Limitations
[31]	Role-based Twitter user's classification.	CNN with multiple classifiers i.e. Decision Tree, SVM, AdaBoost, Gradient Boosting, & Random Forest.	CrowdFlower & Gender Labelled Dataset	89.9	The researchers want to use more features and deploy deep neural networks for enhancing feature fusion.
[32]	Inferring user types as males, females, and organizations.	CNN has been utilized for image learning whereas for classification different techniques are used in this study.	CrowdFlower & ILLAE	85.99 & 78.61	Profile pictures of organizations could not be recognized and female with short hair are identified as male.

		For CrowdFlower Random Forest has achieved the highest accuracy & for ILLAE SVM does best.			
[33]	Gender and Age recognition of Twitter users.	Numerous classification techniques have been tried, and Extra Trees (1,000 classifiers with stop words) has outperformed all other techniques.	CrowdFlo wer	91.4	For gender classification, an ensemble is intended to be applied.

5.3Analysis

This thesis has actually been founded on the technique that has been presented in [34]. The main objective of the thesis is to evaluate the presented methodology on a different dataset than they have used. Seeing the accuracy of the proposed technique ensures that the presented technique is top-notch. Due to the fact that different datasets have been inputted to the proposed system and it has performed really well. Based on the stated performance, it is also evident that the proposed technique does not have any sort of overfitting or underfitting.

As the objective of this thesis is to execute a readily proposed technique in order to assess its performance on a different dataset, this chapter is of high quality. The chapter provides a graph which illustrates the accuracy achieved by the presented system, another graph is used to present the ratio of iterations and loss. This chapter also has a comparison of the results attained by the proposed technique with those achieved by the state-of-the-art techniques applied to the same dataset. The chapter ends with an analysis of the technique.

CHAPTER 6

CONCLUSION AND FUTURE WORK

This is the last chapter of the thesis and is used to conclude the manuscript. It presents the conclusion of the current work and the work that can be done in the future in order to enhance its performance or evaluate it on another level.

6.1Conclusion

Social media platforms and particularly microblogging sites, for instance, Twitter has gradually become a vital part of millions of people around the globe. Apart from interacting and communicating with acquaintances, friends, and family, such services are being used as recommendation services, as sources of real-time news and venues of content sharing.

The experience of a user with microblogging services can be improved quite notably given that information regarding their demographic properties or interests that are personal to a specific user. Information of such sort could be used for recommendations that are personalized or the posts presented to a particular user would be personalized. The topic of interests and events highlighted to a user would be specific and tailored.

Such information of user-profiles for instance name, location, age, gender, and a brief about their interests is more often than not available as far as most social media sites are concerned. However, these details can be incomplete due to the fact that the user does not feel like sharing their details. Or if the information is provided, it may be misleading or imaginary. Moreover, some details may be shared explicitly, for instance, political affiliation. Despite the fact that such information might not be shared in the first place or fake details might be provided. There still are certain ways to predict these attributes.

The rapid growth of social networks has produced an unprecedented amount of user-generated data, which provides an excellent opportunity for text mining. Authorship analysis, an important part of text mining, attempts to learn about the author of the text through subtle variations in the writing styles that occur between gender, age, and social groups. Such information has a variety of applications including advertising and law enforcement. One of the most accessible sources of user-generated data is Twitter, which makes the majority of its user data freely available through its data access API.

This thesis makes use of a convolutional neural network in order to predict the gender of the twitter user. For the stated purpose it uses the CrowdFlower dataset, which can be found on Kaggle (Kaggle 2016) is a project of CrowdFlower (CrowdFlower 2015), including the information of about 20,000 users. Project contributors manually labeled each user by checking the corresponding information, which contains part of the profile metadata, such as display name, screen name, description, link color, etc. There are three labels in the dataset: male, female, and brand. The contributors also provided a confidence score along with the role tag, which is a good indicator of labeling quality.

After some preprocessing the user tweets are fed to the CNN where they are received by the embedding layer. The best-suited architecture of neural network has three 4D convolutional layers. Usually, forward propagation and backward propagation techniques are used as error functions or weight optimization. But in our case, we use the Adaptive moment estimation technique or more commonly known as Adam.

The Softmax regression is a form of logistic regression that normalizes an input value into a vector of values that follows a probability distribution whose total sums up to 1. The output values are between the range [0,1] which is commendable due to the fact that if needed more classes can be adjusted the neural network model.

The mean accuracy that has been achieved by the proposed system is over 97 percent. The stated figure is close to 100 percent and thus the proposed system can be further improved in order to form an automated prediction system and can be made use of for numerous purposes including tailored advertisements and other such campaigns. This work is also applicable to other domains where identifying organizations, for instance, emergency management, is important and where identifying female or male participants is crucial (any campaign directed towards diversity of issues such as gender-based violence).

6.2Future Work

Despite the fact that the accuracy achieved by the proposed system is close to 100 percent, there is still room for improvement and in the future, the accuracy can be increased by trying various experimentations. There are different activation functions that can be tried and their results could be assessed. Various optimization functions can be tried in order to enhance the performance of the system. Moreover, different loss functions can also be exasperated for increasing the accuracy of the system.

All these different combinations can also be tried with a dataset that contains some figures or images of users. With the modern set of tools if both text and images are used to train a network the results can be phenomenal.

This is the concluding chapter of the thesis and presents the conclusion of the study as well as its future work. The conclusion section briefly discusses the architecture of the network as well as some details regarding the dataset. Whereas the future work section contains the different possibilities and options for enhancing the performance of the system.

REFERENCES

[1] Aric Bartle, Jim Zheng. Gender Classification with Deep Learning.

[2] John D. Burger, John Henderson, George Kim, Guido Zarrella. Discriminating Gender on Twitter. Proceedings of the Conference on Empirical Methods in Natural Language Processing, 2011.

[3] Arjun Mukherjee, Bing Liu. Improving Gender Classification of Blog Authors. Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing, 2010.

[4] Ruchita Sarawgi, Kailash Gajulapalli, and Yejin Choi. Gender Attribution: Tracing Stylometric Evidence Beyond Topic and Genre. Conference on Computational Natural Language Learning, 2011

[5] Shlomo Argamona, Moshe Koppel, Jonathan Finec, Anat Rachel Shimoni. Gender, Genre, and Writing Style in Formal Written Texts, Text-Interdisciplinary Journal, 2003

[6] David Bamman, Jacob Einstein, Tyler Schnoebelen. GENDER IN TWITTER: STYLES, STANCES, AND SOCIAL NETWORKS.

[7] Jose Ahirton Batista Lopes Filho , Rodrigo Pasti, Leandro Nunes de Castro. Gender Classification of Twitter Data Based on Textual Meta-Attributes Extraction

[8] J. Kietzmann, K. Hermkens, I. McCarthy and B. Silvestre, "Social media? Get serious! Understanding the functional building blocks of social media", *Business Horizons*, vol. 54, no. 3, pp. 241-251, 2011. Available: https://www.sciencedirect.com/science/article/pii/S0007681311000061. [Accessed 9 May 2019].

[9] A. Kaplan and M. Haenlein, "Users of the world, unite! The challenges and opportunities of Social Media", *Business Horizons*, vol. 53, no. 1, pp. 59-68, 2010. Available: https://www.sciencedirect.com/science/article/pii/S0007681309001232. [Accessed 9 May 2019].

[10] B. Marr, "How Much Data Do We Create Every Day? The Mind-Blowing Stats Everyone Should Read", *Forbes.com*, 2019. [Online]. Available: https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/#2be71ee360ba. [Accessed: 09-May- 2019].

[11] I. Ahmad and I. Ahmad, "How Much Data Is Generated Every Minute? [Infographic]", *Social Media Today*, 2019. [Online]. Available: https://www.socialmediatoday.com/news/how-much-data-is-generated-every-minute-infographic-1/525692/. [Accessed: 10- May- 2019].

[12] A. Vinay, V. Shekhar, J. Rituparna, T. Aggrawal, K. Murthy and S. Natarajan, "Cloud Based Big Data Analytics Framework for Face Recognition in Social Networks Using Machine Learning", *Procedia Computer Science*, vol. 50, pp. 623-630, 2015. Available:

https://www.sciencedirect.com/science/article/pii/S1877050915005967. [Accessed 11 May 2019].

[13] M. Pennacchiotti and A. Popescu, "A Machine Learning Approach to Twitter User Classification", in *Fifth International AAAI Conference on Weblogs and Social Media*, Sunnyvale, USA, 2011, pp. 281-288.

[14] Q. You, J. Luo, H. Jin and J. Yang, "Robust Image Sentiment Analysis Using Progressively Trained and Domain Transferred Deep Networks", in *Twenty-Ninth AAAI Conference on Artificial Intelligence*, 2015.

[15] H. Lin et al., "User-level psychological stress detection from social media using deep neural network", *Proceedings of the ACM International Conference on Multimedia - MM '14*, 2014. Available: https://dl.acm.org/citation.cfm?id=2654945. [Accessed 16 May 2019].

[16] Y. Lv, Y. Duan, W. Kang, Z. Li and F. Wang, "Traffic Flow Prediction With Big Data: A Deep Learning Approach", *IEEE Transactions on Intelligent Transportation Systems*, pp. 1-9, 2014. Available: https://ieeexplore.ieee.org/document/6894591. [Accessed 19 May 2019].

[17] J. Yiny, S. Karimiy, A. Lampertz, M. Camerony, B. Robinsony and R. Powery, "Using Social Media to Enhance Emergency Situation Awareness: Extended Abstract", in *Twenty-Fourth International Joint Conference on Artificial Intelligence (IJCAI 2015)*, Australia, 2015.

[18] F. Greaves, D. Ramirez-Cano, C. Millett, A. Darzi and L. Donaldson, "Harnessing the cloud of patient experience: using social media to detect poor quality healthcare: Table 1", *BMJ Quality & Safety*, vol. 22, no. 3, pp. 251-255, 2013. Available: https://qualitysafety.bmj.com/content/22/3/251. [Accessed 21 June 2019].

[19] S. Daneshvar and D. Inkpen, "Gender Identification in Twitter using N-grams and LSA: Notebook for PAN at CLEF 2018", 2018.

[20] L. Akhtyamova, J. Cardiff and A. Ignatov, "Twitter Author Profiling UsingWord Embeddings and Logistic Regression Notebook for PAN at CLEF 2017", Ireland, 2017.

[21] J. Alowibdi, U. Buy and P. Yu, "Empirical Evaluation of Profile Characteristics for Gender Classification on Twitter", *2013 12th International Conference on Machine Learning and Applications*, 2013. Available: https://ieeexplore.ieee.org/document/6784644/. [Accessed 24 May 2019].

[22] D. Fernández, D. Moctezuma and O. Siordia, "Features combination for gender recognition on Twitter users", in *2016 IEEE International Autumn Meeting on Power, Electronics and Computing (ROPEC 2016).*, 2016.

[23] J. Filho, R. Pasti and L. de Castro, "Gender Classification of Twitter Data Based on Textual Meta-Attributes Extraction", New Advances in Information Systems and Technologies, pp. 1025-1034, 2016. Available: https://link.springer.com/chapter/10.1007/978-3-319-31232-3_97. [Accessed 27 May 2019].

[24] W. Deitrick, Z. Miller, B. Valyou, B. Dickinson, T. Munson and W. Hu, "Gender Identification on Twitter Using the Modified Balanced Winnow", Communications and Network, vol. 04, no. 03, pp. 189-195, 2012. Available: http://file.scirp.org/Html/1-6101209_22061.htm. [Accessed 27 May 2019].

[25] L. Geng, K. Zhang, X. Wei and X. Feng, "Soft Biometrics in Online Social Networks: A Case Study on Twitter User Gender Recognition", 2017 IEEE Winter Applications of Computer Vision Workshops (WACVW), 2017. Available: https://ieeexplore.ieee.org/document/7912201. [Accessed 28 May 2019].

[26] Z. Miller, B. Dickinson and W. Hu, "Gender Prediction on Twitter Using Stream Algorithms with N-Gram Character Features", International Journal of Intelligence Science, vol. 02, no. 04, pp. 143-148, 2012. Available: http://file.scirp.org/Html/4-1680048_24184.htm. [Accessed 27 May 2019]

[27] G. Levi and T. Hassncer, "Age and gender classification using convolutional neural networks", 2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2015. Available: https://ieeexplore.ieee.org/document/7301352. [Accessed 28 May 2019].

[28] J. Burger, J. Henderson, G. Zarrella and G. Zarrella, "Discriminating gender on Twitter", 2011 Conference on Empirical Methods in Natural Language Processing,, pp. 1301-1309, 2011. Available: https://dl.acm.org/citation.cfm?id=2145568. [Accessed 29 May 2019].

[29] M. Vicente, F. Batista and J. Carvalho, "Twitter gender classification using user unstructured information", 2015 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), 2015. Available: https://ieeexplore.ieee.org/document/7338102. [Accessed 29 May 2019].

[30] N. Ljubešić, D. Fišer and T. Erjavec, "Language-independent Gender Prediction on Twitter", Proceedings of the Second Workshop on NLP and Computational Social Science, 2017. Available: https://www.aclweb.org/anthology/papers/W/W17/W17-2901/. [Accessed 30 May 2019].

[31] M. Arroju, A. Hassan and G. Farnadi, "Age, Gender and Personality Recognition using Tweets in a Multilingual Setting", 2015.

[32] S. Bergsma, M. Dredze, B. Van Durme, T. Wilson and D. Yarowsky, "Broadly Improving User Classification via Communication-Based Name and Location Clustering on Twitter", in Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2013, pp. 1010-1019z.

[33] M. Ciot, M. Sonderegger and D. Ruths, "Gender Inference of Twitter Users in Non-English Contexts", in 2013 Conference on Empirical Methods in Natural Language Processing, 2013.

[34] C. Fink, J. Kopecky and M. Morawski, "Inferring Gender from the Content of Tweets: A Region Specific Example", in Sixth International AAAI Conference on Weblogs and Social Media, 2012.

[35] A. Culotta, N. Ravi and J. Cutler, "Predicting the demographics of twitter users from website traffic data", in AAAI'15 Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, 2019, pp. 72-78.

[36] L. De Silva and E. Riloff, "User Type Classification of Tweets with Implications for Event Recognition", Proceedings of the Joint Workshop on Social Dynamics and Personal Attributes in Social Media, 2014. Available: https://www.researchgate.net/publication/301409221_User_Type_Classification_of_Tweets_with_Implications_for_Event_Recognition. [Accessed 4 June 2019].

[37] M. Merler, L. Cao and J. Smith, "You are what you tweet&#x2026;pic! gender prediction based on semantic analysis of social media images", 2015 IEEE International Conference on Multimedia and Expo (ICME), 2015. Available: https://ieeexplore.ieee.org/abstract/document/7177499. [Accessed 4 June 2019].

[38] L. Li, Z. Song, X. Zhang and E. A. Fox, "A Hybrid Model for Role-related User Classification on Twitter", 2018.

[39] H. Karbasian, H. Purohit, R. Handa, A. Malik and A. Johri, "Real-Time Inference of User Types to Assist with more Inclusive and Diverse Social Media Activism Campaigns", Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society - AIES '18, 2018. Available: https://www.researchgate.net/publication/330299876_Real-Time_Inference_of_User_Types_to_Assist_with_more_Inclusive_and_Diverse_Social_Media_Activism_Campaigns. [Accessed 7 June 2019].

[40] M. Kokan and L. Skugor, "Sentiment Analysis of Tweets Using Semantic Reinforced Bag-of-Words Models", in Text Analysis and Retrieval 2017: Course Project Reports (TAR 2017), 2017, pp. 40-43.

[41] Bahdanau, D., Cho, K., Bengio, Y.: Neural machine translation by jointly learning to align and translate. In: Proceedings of the 3rd International Conference on Learning Representations (2014), http://arxiv.org/abs/1409.0473

[42] Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. CoRR abs/1412.6980 (2014), http://arxiv.org/abs/1412.6980

[43] R. Ruizendaal, "Deep Learning #4: Why You Need to Start Using Embedding Layers", *Towards Data Science*, 2019. [Online]. Available: https://towardsdatascience.com/deep-learning-4-embedding-layers-f9a02d55ac12. [Accessed: 08- Jul- 2019].

[44] S. Kim, "A Beginner’s Guide to Convolutional Neural Networks (CNNs)", *Towards Data Science*, 2019. [Online]. Available: https://towardsdatascience.com/a-beginners-guide-to-convolutional-neural-networks-cnns-14649dbddce8. [Accessed: 09- Jul- 2019].

[45] H. Pokharna, "The best explanation of Convolutional Neural Networks on the Internet!", *Medium*, 2019. [Online]. Available: https://medium.com/technologymadeeasy/the-best-explanation-of-convolutional-neural-networks-on-the-internet-fbb8b1ad5df8. [Accessed: 09- Jul- 2019].

[46] Keras, "Merge Layers - Keras Documentation", *Keras.io*, 2019. [Online]. Available: https://keras.io/layers/merge/. [Accessed: 09- Jul- 2019].

[47] MissingLink, "Using the Keras Flatten Operation in CNN Models with Code Examples - ", *Missinglink.ai*, 2019. [Online]. Available: https://missinglink.ai/guides/deep-learning-frameworks/using-keras-flatten-operation-cnn-models-code-examples/. [Accessed: 10- Jul- 2019].

[48] J. Brownlee, "Dropout Regularization in Deep Learning Models With Keras", *Machine Learning Mastery*, 2019. [Online]. Available: https://machinelearningmastery.com/dropout-regularization-deep-learning-models-keras/. [Accessed: 09- Jul- 2019].

[49] J. Brownlee, "Gentle Introduction to the Adam Optimization Algorithm for Deep Learning", *Machine Learning Mastery*, 2019. [Online]. Available: https://machinelearningmastery.com/adam-optimization-algorithm-for-deep-learning/. [Accessed: 08- Jul- 2019].

[50] Ruder, "An overview of gradient descent optimization algorithms", *Sebastian Ruder*, 2019. [Online]. Available: http://ruder.io/optimizing-gradient-descent/. [Accessed: 04- Jul- 2019].

[51]]H. Mahmood, "Softmax Function, Simplified", *Towards Data Science*, 2019. [Online]. Available: https://towardsdatascience.com/softmax-function-simplified-714068bf8156. [Accessed: 07- Jul- 2019].

[52] SuperDataScience, "SuperDataScience", *Superdatascience.com*, 2019. [Online]. Available: https://www.superdatascience.com/blogs/convolutional-neural-networks-cnn-softmax-crossentropy. [Accessed: 07- Jul- 2019].