Author Profiling for Irony Detection based on Natural Language Processing Techniques

A Project Report
Submitted in partial fulfillment of the
requirements for the award of the degree
of
Master of Engineering
in
COMPUTER SCIENCE & ENGINEERING

Under the Supervision of

Mr. Sumit Gupta

Assistant Professor
Department of Computer Science and Engineering
University Institute of Technology
The University of Burdwan

By

Ankan Sinha

Regn. no.- 202020000014 of 2021-22, Roll no.- ME202010008

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
UNIVERSITY INSTITUTE OF TECHNOLOGY
THE UNIVERSITY OF BURDWAN
GOLAPBAG (NORTH), BURDWAN-713104, WEST BENGAL, INDIA

JUNE 2022

UNIVERSITY INSTITUTE OF TECHNOLOGY
THE UNIVERSITY OF BURDWAN

CERTIFICATE OF APPROVAL

*This is to certify that the project entitled "**Author Profiling for Irony Detection based on Natural Language Processing Techniques**" is successfully carried out by **Mr. Ankan Sinha** (Regn. no.- 202020000014 of 2020-21, Roll no.- ME202010008) of University Institute of Technology, The University of Burdwan, Burdwan, West Bengal, India, under my guidance and supervision towards the partial fulfillment of the academic requirements for the award of the degree of Master of Engineering (M.E.) in Computer Science & Engineering (CSE).*

Signed by:

Mr. Sumit Gupta
Project Supervisor & Assistant Professor
Department of Computer Science & Engineering
University Institute of Technology
The University of Burdwan
Burdwan-713104, West Bengal, India

Date:

Place:

UNIVERSITY INSTITUTE OF TECHNOLOGY
THE UNIVERSITY OF BURDWAN

CERTIFICATE

*This is to certify that the project report entitled, "**Author Profiling for Irony Detection based on Natural Language Processing Techniques**" submitted by **Mr. Ankan Sinha** (Regn. no.- 202020000014 of 2020-21, Roll no.- ME202010008) to University Institute of Technology, The University of Burdwan, Burdwan, West Bengal, India, is a record of bona-fide Project work carried out by him towards the partial fulfillment of the academic requirements for the award of the degree of Master of Engineering (M.E.) in Computer Science & Engineering (CSE).*

We wish him every success in his future endeavors.

_____

Mr. Sumit Gupta
Project Supervisor & Assistant Professor
Department of Computer Science & Engineering
University Institute of Technology
The University of Burdwan
Burdwan-713104, West Bengal, India

_____

Dr. Souvik Bhattacharyya
In-Charge & Assistant Professor
Department of Computer Science & Engineering
University Institute of Technology
The University of Burdwan
Burdwan-713104, West Bengal, India

UNIVERSITY INSTITUTE OF TECHNOLOGY
THE UNIVERSITY OF BURDWAN

DECLARATION

I, ***Ankan Sinha*** *(Regn. no.- 202020000014 of 2020-21, Roll no.- ME202010008, hereby declare that the project work in form of this report entitled "****Author Profiling for Irony Detection based on Natural Language Processing Techniques****" submitted to University Institute of Technology, The University of Burdwan, Burdwan, West Bengal, India, towards the partial fulfillment of the academic requirements for the award of the degree of Master of Engineering (M.E.) in Computer Science & Engineering (CSE) is my original work and not submitted for any other award or other degree or any other similar titles or prizes. I also declare that all the details, particulars and information presented in this work have been duly acknowledged under the purview of ethical standards and guidelines.*

_____

Ankan Sinha

Regn. no. 202020000014 of 2021-22

Roll no. ME202010008

Date:

Place:

Acknowledgement:

I have taken efforts in developing this project titled "**Author Profiling for Irony Detection based on Natural Language Processing Techniques**". However, it would not have been possible without the kind support and help of many individuals and my Institution, University Institute of Technology, The University of Burdwan. I would like to extend my sincere thanks to **Dr. Souvik Bhattacharyya**, In- charge of CSE Department, University Institute of Technology, The University of Burdwan and all faculty and staff members of the department.

I am highly indebted to **Mr. Sumit Gupta**, Assistant Professor of CSE Department, University Institute of Technology, The University of Burdwan for the guidance and constant supervision as well as for providing necessary information regarding the project & also for the support in completing the project.

I would like to express my gratitude towards my parents & member of University Institute of Technology, The University of Burdwan for their kind co-operation and encouragement which helped me in completion of this project. I would like to express my special gratitude and thanks to everyone for giving me such attention and time.

_____

Ankan Sinha

Date: **Regn. no. 202020000014 of 2021-22**

Roll no. ME202010008

Place:

Table of Contents:

List of Figures and Table:

Abstract:

The project is focused on profiling ironic authors in Twitter. Special emphasis will be given to those authors that employ irony to spread stereotypes, for instance, towards women or the LGTB community. The goal will be to classify authors as ironic or not depending on their number of tweets with ironic content. Among those authors a subset that employs irony to convey stereotypes will be considered, in order to investigate if state-of-the-art models are able to distinguish also these cases. Therefore, given authors of Twitter together with their tweets, the goal will be to profile those authors that can be considered as ironic. Here in this project, various ML models are used to verify the data provided based on **Homonyms** to detect the Ironical users and distinguished them from the Non-Ironical spreaders, thus have obtained an optimal result which encourage us to proceed further in the near future based on this approach of using homonyms to represent a Ironical situation.

Keywords: **Authorship attribution, Irony, Homonyms, Machine Learning, LR, SVM, Accuracy**

1 INTRODUCTION

Social Network Sites (SNSs) are an ideal place for Internet users to keep in touch, share information abouttheir daily activities and interests, publishing and accessing documents, photos, and videos. SNSs like Facebook, Twitter, YouTube, and Instagram give the ability to create profiles, to have a list of peers to interact with and to post and read what others have posted. It comes as no surprise that, overall, SNSs - together with search engines - are among the most visited websites. Unfortunately, SNSs are also the ideal plaza for proliferation of harmful information. Cyberbullying, sexual predation, self-harm practices incitement is some of the effective results of the dissemination of malicious information on SNSs. Many ofthese attacks are often carried by a single individual, but they can be also managed by groups. The targetof the trolls are often selected victims but, in some circumstances, the hate can be directed towards widegroups of individuals, discriminated for some features, like race or gender. Such campaigns may involve a very large number of haters that are self-excited by hateful discussions, and such hate might end up with physical violence or violent actions.

Unlike topic detection e.g. separating articles from different sports, irony detection is considered as a very difficult task. In the case of classifying sports, each sport (topic) normally has its own vocabulary that is unlikely to appear in any other document than that sport. This vocabulary is often enough for making an accurate classifier. However, irony detection is considered as a very difficult task due to that it appears in various forms and in all kinds of contexts. Irony often has ambiguous interpretations, and it often expresses the contrary to what is literally being said [Butler, 1953]. Another reason why irony could be more difficult to detect in written text than in spoken conversation is e.g., due to that information contained in the tone of the voice or facial expressions that can be important for understanding irony are lost. Yet another theory is that it takes longer to write on a computer than talking face-to-face, and that people use that extra time to write more complex sarcasm than what they use normally.

As we are entering the era of Big data there is an increasing amount of data being stored and processed every day, containing everything from customers shopping behavior to how different medicines act on patients. Handling this amount of data calls for an automated tool for data analysis, which is what Machine Learning provides.

1.1 Why is irony detection interesting?

Except for being an interesting field for human-computer interaction irony detection is also an important area in sentiment analysis and opinion mining. Understanding irony in text would enable a more accurate picture of what is requested [Carvalho et al., 2009]. In medical care sarcasm detection could work as detection for brain injuries in an early stage. In research done at the University College of London they were able to show that people suffering brain injuries had impaired sarcasm comprehension compared to the control group [Channon et al., 2004]. Another important area is for security reasons to evaluate whether a threat is real or not. The consequences could be devastating if some ironic posts were to be taken seriously.

E.g., the humorous twitter account dialyses tweeted" And 12 bombers depart now ... to Lithuania!" after it turned out that Lithuania gave Russia zero points for its performance in the Eurovision Song Contest 2015. U.S. secret service also said that they are starting to work on a sarcasm detector for twitter. Even though experts in the area express skepticism towards the idea, it is considerable 4 that at least two individuals have been falsely arrested because of sarcastic posts during the past five year

2 LITERATURE SURVEY

• In Paper by Watanabe Et al. [1], the offensive and hate speech from twitter is being detected based on the writing pattern and unigram along with sentimental features. They have classified their dataset into 3 parts of which the 1st and 2nd one is created manually with 14K tweets and is classified into hateful, offensive, and clean and the 3rd one is classified into sexism and racism. Theyused Features extraction and then Parametric optimization to obtain the result based on the binary and ternary classification.

• In paper by Jain Et al. [2], Offensive and hate speech in English and Spanish from Twitter is being detected based on KNN Classifier, Multinomial Naïve Bayes, Logistic regression and linear SVM. The approach is based on assumption of an ID as a Hate speech spreaderand thus detecting the result. The Experimental result for the one is obtained by using machine learning and deep learning process.

• In paper by Badjatiya Et al. [3], Offensive and hate speech from Twitter is being detected based on multiple deep learning architectures. The process detects only Tweets as racist, sexist, and neutral. Here CNN is used for hate speech and sentiment analysis and LSTM is used to capture long range dependency on Tweets. And after the final evaluation the result thus obtained is F1 Score as 0.930.

• In Paper by Davidson Et al. [4], Hate speech from Twitter is being detected by using crowd sourced lexicon. But here Sexist Tweets are generally classified as Offensive, Tweets without explicit hate keywords are also more difficult to classify. Here the datas are testedwith Logistic regression, Naïve Bayes, Random Forest etc. where Each model is tested with 5-fild cross validation which results in Logistic Regression and Linear SVM to be better and the logistic regression with L2 regularization is used as final model to obtain the best result.

• In Paper by Ashraf Et al. [5], Detection of hate speech from YouTube video comments is being done. A Total of 400 Videos have beengathered using YouTube API for performing the experiment. The classified data is being evaluated using SVM, LR and KNN, thus in result the obtained accuracy for SVM is greater than the other 2 models.

- In Paper by Jens Lemmens Et al. [12], an approach for the detection of sarcasm in Reddit and Twitter responses in the context of The Second Workshop on Figurative Language Processing held in conjunction with ACL 20201 . The ensemble is trained on the predicted sarcasm probabilities of four component models and on additional features, such as the sentiment of the comment, its length, and source (Reddit or Twitter) in order to learn which of the component models is the most reliable for which input. The component models consist of an LSTM with hashtag and emoji representations; a CNN-LSTM with casing, stop word, punctuation, and sentiment representations; an MLP based on Infersent embeddings; and an SVM trained on stylometric and emotion-based features. All component models use the two conversational turns preceding the response as context, except for the SVM, which only uses features extracted from the response.

- In paper by Zhang Zuping, Dmien and Long[10], Hate Speech Detection is being done only based on Lexicon. A sentence-level test annotation was manually carried out on a representative sample categorizing the hate corpus into 3 different categories. Then, usingsubjectivity analysis, objective sentences were separated from subjective sentences and removed from the corpora. A lexicon was created from semantic, hate and theme-based features and used in creating a rule-based classifier for hate speech detection.

- In Paper by Erik Forslid, Niklas Wikén[11], irony and sarcasm detection and also includes the design and programming of a machine learning model that classifies text as sarcastic or non-sarcastic. This is done with supervised learning. Two different data set where used, one with Amazon reviews and one from Twitter. An accuracy of 87% was obtained on the Amazon data with the Support Vector Machine. For the Twitter data was an accuracy of 71% obtained with the Adaboost classifier was used. The thesis is done in collaboration with Gavagai AB, which is company working with Big-data text with expertise in semantic analysis and opinion miniIn Paper by Sureka Et al. [6], Cyber hate detection from YouTube is being detected based on User subscriptions. Total 75 Videos have been gathered using YouTube API for the experiment. Linguistic analysis of comments and social network analysis is being done to obtain the experimental result based on a propped framework.

- In Paper by Del Vigna Et al. [7], Hate Speech detection from Facebook posts and comments is being done using SVM LSTM Lexicon Classifier. A versatile Facebook crawler have been developed by them, which exploits the Graph API to retrieve the content of the comments to Facebook posts. The experimented result is obtained by using the classifiers over the dataset.

- In Paper by Ben-David and Fernandez [8], Hate Speech and Covert Discrimination on the Facebook Pages of Extreme Right Political Parties in Spain is being detected by using Facebook API Textual analysis Image and Link analysis Network Analysis. The study analytically distinguished between instances of overt hate speech, which could be regarded as violating the platform's community andbetween covert practices, which were not addressed by the platform's community standards but nonetheless discriminated through the interaction between users and the platform's technological affordances.

- In Paper by Ahmed Et al. [9], Detection and classification on of social media-based extremist affiliations using sentiment analysis techniques is being done based on twitter's tweets using LSTM, CNN and ML and DL. This experiment is basically done to detect the extremis posts wrt. ISIS, Bombs etc. The experimental accuracy is found to be 90% in this cases

Table 1: Tabular Comparison of Prominent related works on Author Profiling.

Sl. No.	Author	Objective	Limitations	Features used	Dataset used	Technique	Reported Results
1	Ø Hajime Watanabe, Ø Mondher Bouazizi, Ø Tomoaki Ohtsuki	Offensive and hate speech from Twitter is being detected based on writing pattern and unigrams along with sentimental features	This approach is based on a pre- defined training set and only for Twitter. The tweets in the dataset which is categorized as "Neither", many- times found to be belonging to both Hateful and Offensive, when checked manually	Writing pattern Unigram, Bigram Sentimental Features, Semantic Features	Ø 1st dataset contains more than 14K tweets that have been manually classified into Hateful, Offensive, and clean. Ø 2nd dataset is also classified in the same. Ø 3rd dataset is also tweets but classified into Sexism, Racism.	Ø Test set and a validation set was created from the datasets. Ø Parametric Optimization	Classification was done using toolkit weka. Accuracy = 87.4% on binary classification. Accuracy = 78.4% on ternary classification

Table 1 (Contd.): Tabular Comparison of Prominent related works on Author Profiling.

Sl. No.	Author	Objective	Limitations	Features used	Dataset used	Technique	Reported Results
2	Ø Rakshita Jain Ø Devanshi Goel Ø Prashant Sahu Ø Abhinav Kumar Ø Jyoti Prakash Singh	Offensive and hate speech in English and Spanish from Twitter is being detected based on KNN Classifier, Multinomial Naïve Bayes, Logistic regression and linear SVM The approach is based on assumption of an ID as a Hate speech spreader and thus detecting the result. The LSTM, Bi- LSTM and BERT model was not able to detect the hate speech spreader correctly thus semantics of the texts were not captured.	Ø Lexicon Ø TF-IDF Vector Ø Bag of Words Vector	PAN21-Profiling-Hate-Speech-Spreader-on-Twitter data provided by J.Bevendorff and team.in their paper Overview of PAN2021: Authorship Verification, Profiling Hate Speech Spreader on Twitter, and Style Change Detection, in 12th International conference of CLEF, Springer, 2021	Ø KNN Classifier, Multinomial Naïve Bayes, Ø Logistic regression Linear SVM LSTM and bi-LSTM and BERT (for identification of hate speech spreader) Ø Machine learning and deep learning Logistic regression Ø Random forest, Ø SVMs, GBDT, DNN	**Accuracy:** 66% for English dataset 80% for Spanish Dataset Average: 73% (Result was obtained using Naïve Bayes Classifier with an n-gram range of (1,1) for English and (1,3) for Spanish)

Table 1 (Contd.): Tabular Comparison of Prominent related works on Author Profiling.

Sl. No.	Author	Objective	Limitations	Features used	Dataset used	Technique	Reported Results
3	Ø Pinkesh Badjatiya Ø Shashank Gupta Ø Manish Gupta Ø Vasudeva Varma	Offensive and hate speech from Twitter is being detected based on multiple deep learning architectures.	The process detects only Tweets as racist, sexist and neutral. Images and Videos are not included.		Z.Waseem and D.Hovy, Hateful Symbols or Hateful people? Predictive Features for hate speech Detection on Twitter. In NAACL-HLT, Pages 88-93, 2016	Ø Fast Text, CNN, LSTM.	DNN+GBDT Classifier: LSTM + Random Embedding + GBTD: Prec=Recal =F1=0.930

Table 1 (Contd.): Tabular Comparison of Prominent related works on Author Profiling.

Sl. No.	Author	Objective	Limitations	Features used	Dataset used	Technique	Reported Results
4	Ø Thomas Davidson Ø Dana Warmsley Ø Michael Macy Ø Ingmar Weber	Hate speech from Twitter is being detected by using crowd-sourced lexicon	Sexist Tweets are generally classified as Offensive, Tweets without explicit hate keywords are also more difficult to classify	Ø Bi gram, Ø Un igram Ø Tri gram Ø Se ntiment Lexicon	Random 25K tweets using the term for lexicon by using Twitter API and had them manually coded by CrowdFlower (CF)	Naïve Bayes, Random Forest etc. Logistic Regression and Linear SVM to be better. Ø Weighted by TF-IDF, Ø NLTK for capturing info about syntactic structure and to construct Penn POS.	Precision: 0.91 Recall: 0.90 F1 Score: 0.90 Almost 40% hate speech are misclassified. Precision and Recall score for Hate classes are 0.44 and 0.61
5	Ø Noman Ashraf Ø Abid Rafiq Ø Sabur Butt Ø Hafiz Muhammad Faisal Shehzad	Detection of hate speech from YouTube video comments	Video content is not being detected. Based on religion classification only	Ø Lexicon	Total 400 Videos have been gathered using YouTube API	Weka is used for the experiment Ø SVM Ø LR Ø KNN	Classification accuracy: SVM: 82.6% LR: 82.5% KNN: 80.06%

Table 1 (Contd.): Tabular Comparison of Prominent related works on Author Profiling.

Sl. No.	Author	Objective	Limitations	Features used	Dataset used	Technique	Reported Results
6	Ø Ashish Sureka Ponnurangam Kumaraguru Atul Goyal Ø Sidhart Chhabra	Detection of cyber hate on YouTube	Only hate and extremist videos and users are being detected.	Linguistic analysis of comments Social network analysis	Total 75 Videos have been gathered using YouTube API	A Predefined Framework. Dataset has been seeded using YouTube API LIWC LDA Ego-centric Network Graph around central nodes Subscription, Favorite relationship network	Average Precision: 0.88

Table 1 (Contd.): Tabular Comparison of Prominent related works on Author Profiling.

Sl. No.	Author	Objective	Limitations	Features used	Dataset used	Technique	Reported Results
7	Ø Fabio Del Vigna Ø Andrea Cimino Ø Felice Dell'Orletta Ø Marinella Petrocchi Ø Maurizio Tesconi	Hate Speech detection from facebook posts and comments.	Images posted on Facebook are not included. Only done for Italian language	Lexicon Classifier	A versatile Facebook crawler have been developed, which exploits the Graph API to retrieve the content of the comments to Facebook posts.	The Crawler was used to collect comments from various posts SVM LSTM	Accuracy: SVM: 80.60% LSTM: 79.81%
8	Ø Anat Ben-David Ø Ariadna Matamoros Fernandez	Hate Speech and Covert Discrimination on the Facebook Pages of Extreme-Right Political Parties in Spain	Only Political parties' pages and users related is being covered. It cannot be applied rapidly to the social media platform	Facebook API Lexicon Semantic Syntactic	Software tool Netvizz to retrieve data about content and interactions on specific Facebook pages in a manner similar to Facebook's own collection	Textual analysis Image and Link analysis Network Analysis	The study analytically distinguished between instances of overt hate speech, which could be regarded as violating the platform's community and between covert practices,

Table 1 (Contd.): Tabular Comparison of Prominent related works on Author Profiling.

Sl. No.	Author	Objective	Limitations	Features used	Dataset used	Technique	Reported Results
					of data via its algorithms. The data set included data retrieved from seven Facebook pages of the following Spanish political parties: España 2000 and PxC, two political parties with relevant seats in the municipal assemblies; MSR and AES, two parties with minor local representation; and the three		which were not addressed by the platform's community standards but nonetheless discriminated through the interaction between users and the platform's technological affordances.

Table 1 (Contd.): Tabular Comparison of Prominent related works on Author Profiling.

Sl. No.	Author	Objective	Limitations	Features used	Dataset used	Technique	Reported Results
					parties that constitute the coalition La España en Marcha: FE- La Falange, Nudo Patriota Español (NPe), and DN.6 The Facebook page of the majoritarian party PP was also included in this investigation to compare whether there were similarities in instances of overt hate speech and covert discriminatory		

Table 1 (Contd.): Tabular Comparison of Prominent related works on Author Profiling.

Sl. No.	Author	Objective	Limitations	Features used	Dataset used	Technique	Reported Results
					patterns between the pages of the extreme-right political parties and the governin g party. For NPe, FE-La Falange, España 2000, and PP, the data set covered almost five years of activity and 3 years of interactio n for later Face book joiners.		

Table 1 (Contd.): Tabular Comparison of Prominent related works on Author Profiling.

Sl. No.	Author	Objective	Limitations	Features used	Dataset used	Technique	Reported Results
9	Ø Shakeel Ahmad Ø Muhammad Zubair Asghar Ø Fahad M. Alotaibi Ø Irfanullah Awan	Detection and classifcation of social media-based extremist afliations using sentiment analysis techniques	Lack of an automated method for crawling, cleaning and storing Twitter content. Lack of considering visual and social context features for obtaining more robust results,	Lexicon Semantic	Twitter streaming API is used to scrap tweets containing one or more extremis related keywords (ISIS, bomb, suicide etc.) Dark Web Forum	Users' tweet collection LSTM CNN ML and DL Classification with respect to extremist and no extremist classes using LSTM+CNN model and other ML and DL classifiers	Accuracy: 90% Precision: 86% Recall: 84% F1 Score: 90%

Table 1 (Contd.): Tabular Comparison of Prominent related works on Author Profiling.

Sl. No.	Author	Objective	Limitations	Features used	Dataset used	Technique	Reported Results
10	Ø Njagi Dennis Gitari Zhang Zuping Ø Hanyurwimfura Damien Ø Jun Long	Lexicon-based Approach for Hate Speech Detection	Result is less accurate, No visual content is being included	Lexicon Semantic Semantic+ hate+theme-based	For main source, a diverse date of a total of 100 blog postings (documents) from 10 different websites, 10 for each site, from a list provided in the Hate Directory (directory compiled by Raymond Franklin of sites that are generally offensive) has been checked. The secondary source website consists of largely one paragraph	ML. Subjectivity analysis.	1st. Corpus: Precision: 65.32 Recall: 64.92 F Score: 65.12 2nd. Corpus: Precision: 63.78 Recall: 64.00 F Score: 63.89

Table 1 (Contd.): Tabular Comparison of Prominent related works on Author Profiling.

Sl. No.	Author	Objective	Limitations	Features used	Dataset used	Technique	Reported Results
					snippets of quotes relating to the Israel-Palestinian conflict		

3 NATURAL LANGUAGE PROCESSING

3.1 Overview:

Natural language processing (NLP) refers to the branch of computer science—and more specifically, the branch of artificial intelligence or AI—concerned with giving computers the ability to understand text and spoken words in muchthe same way human beings can.

NLP combines computational linguistics—rule-based modeling of human language—with statistical, machine learning, and deep learning models. Together, these technologies enable computers to process human language in the form of textor voice data and to ‘understand’ its full meaning, complete with the speaker or writer’s intent and sentiment.

NLP drives computer programs that translate text from one language to another, respond to spoken commands, and summarize large volumes of text rapidly—even in real time. There’s a good chance you’ve interacted with NLP in the formof voice-operated GPS systems, digital assistants, speech-to-text dictation software, customer service chatbots, and otherconsumer conveniences. But NLP also plays a growing role in enterprise solutions that help streamline business operations, increase employee productivity, and simplify mission-critical business processes.

3.2 Uses of NLP:

Several NLP tasks break down human text and voice data in ways that help the computer make setoff what it'singesting. Some of these tasks include the following:

I. **Speech recognition**, also called speech-to-text, is the task of reliably converting voice data into text data. Speech recognition is required for any application that follows voice commands or answers spoken questions. What makes speech recognition especially challenging is the way people talk—quickly, slurring words together,with varying emphasis and intonation, in different accents, and often using incorrect grammar.

II. **Part of speech tagging**, also called grammatical tagging, is the process of determining the part of speech of a particular word or piece of text based on its use and context. Part of speech identifies ‘make’ as a verb in ‘I can make a paper plane,’ and as a noun in ‘What make of car do you own?’

III. **Word sense disambiguation** is the selection of the meaning of a word with multiple meanings through a process of semantic analysis that determine the word that makes the most sense in the given context. Forexample, word sense disambiguation helps distinguish the meaning of the verb 'make' in ‘make the grade’(achieve) vs. ‘make a bet’ (place).

IV. **Named entity recognition,** or NEM, identifies words or phrases as useful entities. NEM identifies ‘Kentucky’ asa location or ‘Fred’ as a man's name.

V. **Co-reference resolution** is the task of identifying if and when two words refer to the same entity. The most common example is determining the person or object to which a certain pronoun refers (e.g., ‘she’ = ‘Mary’), butit can also involve identifying a metaphor or an idiom in the text (e.g., an instance in which 'bear' isn't an animal but a large hairy person).

VI. **Sentiment analysis** attempts to extract subjective qualities—attitudes, emotions, sarcasm, confusion,suspicion—from text.

Natural language generation is sometimes described as the opposite of speech recognition or speech-to-text;it's the task of putting structured information into human language.

4 ONLINE COMMUNITIES

4.1 Overview:

An online community is simply a group of people coming together for a common purpose, interest, or vision, and doing so via theinternet. Online communities typically use chat rooms, mailing lists, and forums as their primary mode of interaction.

Online communities tend to form around hobbies, special interests, creators, influencers, or even companies and their products. They take many forms and express themselves in a multitude of ways. While many virtual communities stay online-only, some havein-person events and elements as well.

The thing that differentiates online vs. local communities is the primary mode of engagement. Local communities focus membership around a geographic area, where anyone from any location can participate in a community that's based only.

4.2 Types of Online Communities:

There are three types of online communities, each with a unique purpose:

- **Support communities**: A place for users to request and provide help on a specific subject, such as an auto-repair community where people can ask for maintenance and repair help.

- **Discussion communities**: A place for users to discuss a common interest, such as a community where people can talk about their favorite TV show, sports team, or a hobby like woodworking. Ex.: Facebook, Orkut, Twitter.

- **Action communities**: A place for users to plan and work towards a common goal together, such as a community focused on organizing fundraisers for charity causes. Ex.: Slack, Google Workspace

5

AUTHOR PROFILING & IRONY DETECTION

5.1 Overview:

Author profiling is the analysis of a given set of texts to uncover various characteristics of the author based on stylistic- and content-based features, or to identify the author. Characteristics analyzed commonly include age and gender, though more recent studies have looked at other characteristics like personality traits and occupation

5.2 Definition of Irony:

Irony is a sophisticated form of language use that acknowledges a gap between the intended meaning and the literal meaning of the words. Even though it is a widely studied linguistic phenomenon no clear definition seems to exist [Filatova, 2012]. Irony can be divided into two broad categories: situational and verbal irony. The former one is e.g. a cigarette company having non-smoking signs in the lobby. The latter one, which is considered in this master thesis, is most commonly defined as ”saying the opposite of what you mean” [Butler, 1953] where the difference between the saying and meaning is supposed to be clear. Other definitions states that it is any form of negation without any markers [Giora, 1995], another one says that irony violates the maxim of not saying what you believe is false [Grice, 1975]. Yet another definition is that an utterance has to be echoic to be judged as ironic [Sperber and Wilson, 1995].

1. *”thank you Janet Jackson for yet another year of Super Bowl classic rock!”*
2. *”He’s with his other woman: XBox 360. It’s 4:30 fool. Sure I can sleep through the gun- fire”*
3. *”WOW I feel your interest...?”*
4. *”[I] Love The Cover”*

These examples demonstrate different situations of irony and sarcasm in different texts.
Example (1), (2) and (3) are tweets from Twitter and example (4) is taken from an Amazon review.

In example (1) it refers to a very poor performance at the super bowl and refers to the disgraceful performance of Janet Jackson the year before.

Example (2) consists of three sentences that are sarcastic on their own, and when combining them the sarcasm becomes obvious.

In example (3) an indication of irony is that" WOW" is spelled with capital letters in combination with the ellipsis at the end of the sentence.

The example from Amazon (4) could just as well be taken from a positive book review.

However, this is taken from a review entitled" Do not judge the book on its cover", and with this title it reveals a sarcastic tone.

To summarize, any linguistic, approach to capture irony theoretically has, rather than come with a final solution, given different perspectives to the phenomena [Sperber and Wilson, 1995]. It is obvious that there is no simple rule or algorithm that can capture irony. In this study we will look at several aspects (sentiments, vocabulary, linguistics, punctuations etc.) that can play a part in forming ironic content.

5.3 Homonyms:

Homonyms are words that are spelled the same and sound the same but have different meanings. The word homonym comes from the prefix *homo-* which means "the same," and the suffix *-nym*, which means "name." Therefore, a homonym is a word that has at least two different meanings, even though all uses look and sound exactly alike.

A simple example of a homonym is the word **pen**. This can mean both "a holding area for animals" and "a writing instrument." Another example is **book**, which can mean "something to read" or "the act of making a reservation." In both cases, the sound and spelling are the same; only the definition changes.

Note that some homonyms have more than two meanings (for example, "tender" can also mean sensitive, easily chewed, or even refer to chicken strips.

Table-2: Example of homonym

Homonym	Meaning 1	Meaning 2
address	to speak to	location
air	oxygen	a lilting tune or voice
arm	body part	division of a company
band	a musical group	a ring
bark	a tree's out layer	the sound a dog makes

5.4 Part-Of-Speech (POS)-tagging:

Another tool that can be used for feature extraction is a POS-tagger. It is one of many interesting tools in stylometry that can be used in Natural language processing. POS taggers assign each token its word class, i.e. tags the token with its part of speech. A benefit of using a POS-tagger is to distinguish homonyms, which is shown in the example below.

1. Put/VB it/PRP back/RB.
2. I/PRP hurt/VBP my/PRP$ back/NN.

Here back is tagged as an adverb (RB) in the first sentence and as a noun (NN) in the second. The POS-tagger used for this example is based on a simple rule-based tagger developed by Eric Brill in 1992. Brill's tagger algorithm has two steps. In the first step, each word is assigned a tag estimated by a large manually tagged corpus. In addition, if new words appear that do not exists in the tagged corpus, it is assigned according to the assumptions: If the first letter is capital then it is assigned a noun tag. Otherwise it is assigned a tag of the class with words that most commonly end with the same three letters. In the second step it examines the tags using a set of rules that examine the order of the tags [Brill, 1992].

5.5. n-grams:

Patterns can be created by concatenating adjacent tokens into n-grams where n = {1, 2, 3...}. For the simple case where n is equal to one is also called unigram. An example of how the n-grams are constructed is explained with the example sentence:" Nicklas likes his new backpack", in Table. By using n-grams it may be possible to capture how a word tend to appear in text with respect to other words. Usually n-grams are never longer than n = 3. Greater values are likely to create" too" complex patterns that rarely match.

Table-3: Example of n-gram

n-gram	Sequence	Length
1-gram:	"Niklas", "likes", "his", "new", "backpack"	5
2-gram:	"Niklas likes", "likes his", "his new", "new backpack"	4
3-gram:	"Niklas likes his", "likes his new", "his new backpack"	3

5.6 Feature selection:

The goal with feature selection is to select a subset to use for classification. In text data there will always be features that are irrelevant and redundant. The redundant features are those that do not contribute anything or very little in distinguishing the classes from each other. By doing this the dimensionality will be reduced so the amount of data to process is less and this will save time when performing the classification. Another benefit, for some classification algorithms, is that we lower the risk of overfitting the data.

5.7 Machine Learning:

Machine learning (ML) is the study of computer algorithms that can improve automatically through experience and by the use of data. It is seen as a part of artificial intelligence. Machine learning algorithms build a model based on sample data,known as training data, in order to make predictions or decisions without being explicitly programmed to do so. Machine learning algorithms are used in a wide variety of applications, such as in medicine, email filtering, speech recognition, and computer vision, where it is difficult or unfeasible to develop conventional algorithms to perform the needed tasks.

5.8 Deep Learning:

Deep learning is a machine learning technique that teaches computers to do what comes naturally to humans: learn by example. Deep learning is a key technology behind driverless cars, enabling them to recognize a stop sign, or to distinguish a pedestrian from a lamppost. It is the key to voice control in consumer devices like phones, tablets, TVs, andhands-free speakers. Deep learning is getting lots of attention lately and for good reason. It's achieving results that were not possible before.

In deep learning, a computer model learns to perform classification tasks directly from images, text, or sound. Deeplearning models can achieve state-of-the-art accuracy, sometimes exceeding human-level performance. Models aretrained by using a large set of labeled data and neural network architectures that contain many layers.

5.9 CNN:

In deep learning, a convolutional neural network (CNN/ConvNet) is a class of deep neural networks, most applied to analyze visual imagery. Now when we think of a neural network, we think about matrix multiplications but that is not thecase with ConvNet. It uses a special technique called Convolution. Now in mathematics convolution is a mathematical operation on two functions that produces a third function that expresses how the shape of one is modified by the other.

5.10 LSTM:

Long short-term memory (LSTM) is an artificial recurrent neural network (RNN) architecture used in the field of deep learning. Unlike standard feedforward neural networks, LSTM has feedback connections. It can process not only singledata points (such as images), but also entire sequences of data (such as speech or video). For example, LSTM is applicable to tasks such as unsegmented, connected handwriting recognition, speech recognition and anomaly detectionin network traffic or IDSs (intrusion detection systems).

5.11 Linear Regression:

Linear regression is one of the easiest and most popular Machine Learning algorithms. It is a statistical method that is usedfor predictive analysis. Linear regression makes predictions for continuous/real or numeric variables such as sales, salary,age, product price, etc.

Linear regression algorithm shows a linear relationship between a dependent (y) and one or more independent (y) variables,hence called as linear regression. Since linear regression shows the linear relationship, which means it finds how the valueof the dependent variable is changing according to the value of the independent variable.

The linear regression model provides a sloped straight line representing the relationship between the variables.

5.12 KNN:

The k-nearest neighbors' algorithm (k-NN) is a non-parametric classification method first developed by Evelyn Fix and Joseph Hodges in 1951, and later expanded by Thomas Cover. It is used for classification and regression. In both cases,

the input consists of the k closest training examples in a data set. The output depends on whether k-NN is used forclassification or regression:

In k-NN classification, the output is a class membership. An object is classified by a plurality vote of its neighbors, with theobject being assigned to the class most common among its k nearest neighbors (k is a positive integer, typically small). If k = 1, then the object is simply assigned to the class of that single nearest neighbor.

In k-NN regression, the output is the property value for the object. This value is the average of the values of k nearestneighbors.

k-NN is a type of classification where the function is only approximated locally, and all computation is deferred until function evaluation. Since this algorithm relies on distance for classification, if the features represent different physicalunits or come in vastly different scales then normalizing the training data can improve its accuracy dramatically.

5.13 SVM:

Support Vector Machine (SVM) is a supervised machine learning algorithm used for both classification and regression. Though we say regression problems as well its best suited for classification. The objective of SVM algorithm is to find a hyperplane in an N-dimensional space that distinctly classifies the data points. The dimension of the hyperplane dependsupon the number of features. If the number of input features is two, then the hyperplane is just a line. If the number of input features is three, then the hyperplane becomes a 2-D plane. It becomes difficult to imagine when the number of features exceeds three.

The advantages of support vector machines are:

- Effective in high dimensional spaces.
- Still effective in cases where number of dimensions is greater than the number of samples.

5.14 BERT:

BERT is an open-source machine learning framework for natural language processing (NLP). BERT is designed to helpcomputers understand the meaning of ambiguous language in text by using surrounding text to establish context. The BERT framework was pre-trained using text from Wikipedia and can be fine-tuned with question-and-answer datasets.

BERT, which stands for Bidirectional Encoder Representations from Transformers, is based on Transformers, a deep learning model in which every output element is connected to every input element, and the weightings between them aredynamically calculated based upon their connection.

5.15 Lexicon:

A lexicon is the vocabulary of a language or branch of knowledge (such as nautical or medical). In linguistics, a lexicon is a language's inventory of lexemes. The word lexicon derives from Greek word λεξικόν (lexikon), neuter of λεξικός (lexikos)meaning 'of or for words'.

Linguistic theories generally regard human languages as consisting of two parts: a lexicon, essentially a catalogue of a language's words (its word-stock); and a grammar, a system of rules which allow for the combination of those words into meaningful sentences. The lexicon is also thought to include bound

morphemes, which cannot stand alone as words (such as most affixes). In some analyses, compound words and certain classes of idiomatic expressions, collocations andother phrases are also considered to be part of the lexicon. Dictionaries represent attempts at listing, in alphabetical order,the lexicon of a given language; usually, however, bound morphemes are not included.

5.16 Semantic Features:

Semantics is the study of meaning in language. It can be applied to entire texts or to single words. For example, "destination" and "last stop" technically mean the same thing, but students of semantics analyze their subtle shades ofmeaning.

In linguistics, semantics is the subfield that studies meaning. Semantics can address meaning at the levels of words, phrases, sentences, or larger units of discourse. Two of the fundamental issues in the field of semantics are that of compositional semantics (which pertains on how smaller parts, like words, combine and interact to form the meaning oflarger expressions such as sentences) and lexical semantics (the nature of the meaning of words).

5.17 Syntactic Features:

Syntax (/ˈsɪntæks/) is the study of how words and morphemes combine to form larger units such as phrases and sentences. Central concerns of syntax include word order, grammatical relations, hierarchical sentence structure(constituency), [3] agreement, the nature of crosslinguistic variation, and the relationship between form and meaning.There are numerous approaches to syntax which differ in their central assumptions and goals.

Syntactic features are formal properties of syntactic objects which determine how they behave with respect to syntactic constraints and operations (such as selection, licensing, agreement, and movement). Syntactic features can be contrasted with properties which are purely phonological, morphological, or semantic, but many features are relevant bothto syntax and morphology, or to syntax and semantics, or to all three components.

5.18 Textual Analysis:

Textual analysis is a methodology that involves understanding language, symbols, and/or pictures present in texts to gaininformation regarding how people make sense of and communicate life and life experiences. Visual, written, or spoken messages provide cues to ways through which communication may be understood. Often the messages are understood as influenced by and reflective of larger social structures. For example, messages reflect and/or may challenge historical,cultural, political, ethical contexts for which they exist. Therefore, the analyst must understand the broader social structures that influence the messages present in the text under investigation.

5.19 Network Analysis:

Network Analysis methods is a group of special analytical methods that are used in case where it is necessary to analyzeand optimize a network of interconnected and related elements that have some connection between one another.

The network analysis methods are used in project management where the elements are key activities of the project in themutual time relation. Another possibility of their use is in the field of logistics and transportation, where the elements represent the center, and the dependencies are spatial (also figuratively temporal). The network analysis methods focus on calculating or **critical path optimizing** between the elements.

The network analysis methods are related to the concept of **network diagram**, which is a view of the project as a diagramwhich expresses various links between the project activities. The network diagrams and network analysis methods are based on the Graph Theory.

5.20 NLTK:

NLTK is a leading platform for building Python programs to work with human language data. It provides easy-to-use interfaces to over 50 corpora and lexical resources such as WordNet, along with a suite of text processing libraries for classification, tokenization, stemming, tagging, parsing, and semantic reasoning, wrappers for industrial-strength NLPlibraries, and an active discussion forum.

6

PROPOSED WORK

6.1 Overview:

We have built our workflow for the data collected and thus create our own way to implement our model.

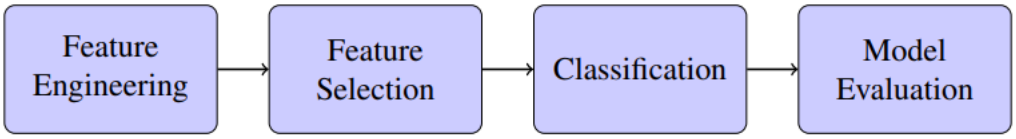

Fig 1: Text Categorization Pipeline

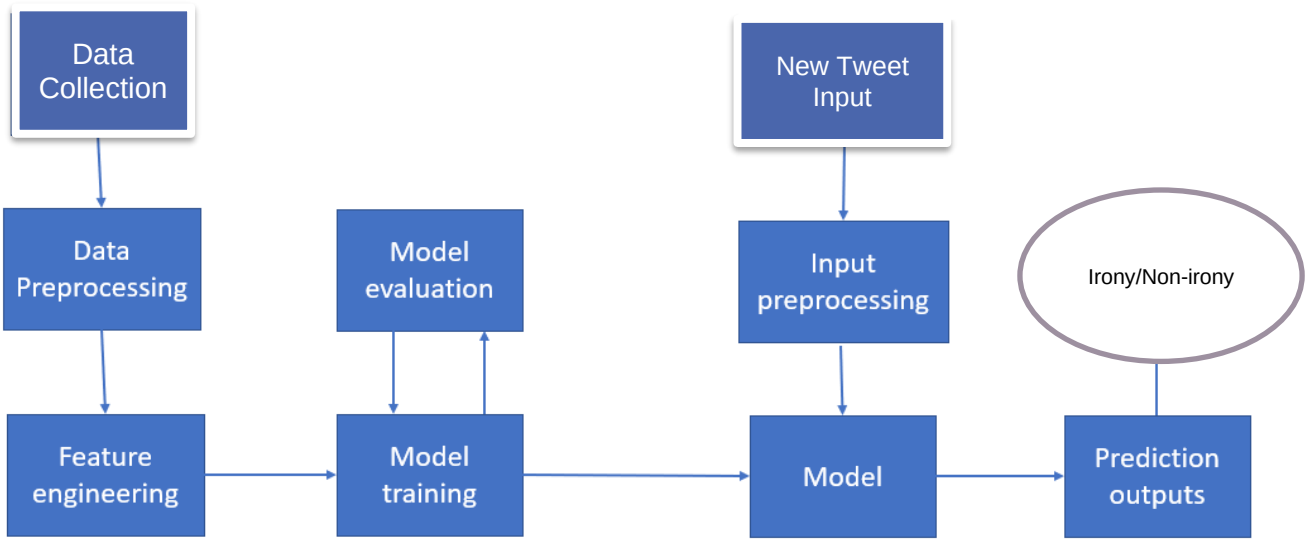

Fig 2: Workflow of our proposed system

6.2 Methodology

The different phases of the proposed methodology for Irony detection using NLP technique are described below.

6.2.1 Data Collection:

Twitter is a micro-blog that allows broadcasting of posts consisting of up to 140 characters and a post is normally referred to as a Tweet. It is possible to write directly to other user by mentioning the other user's username prefixed with a @ character. That tweet will then become visible also to the mentioned users' followers. A tweet may also contain a hashtag prefixing a string of characters with the # character. Users may browse for hashtags; they are therefore normally used for putting a topic to the tweet or commenting on an event etc. Other type of data that are normal in tweets are URLs and pictures.

We downloaded the data set from PAN Clef. PAN is a series of scientific events and shared tasks on digital text forensics and stylometry. The data was a set of 80,000 tweets of 400 users and a truth file which indicates 20 Users to be Ironical and 20 to be non-Ironical. Based on that we need to detect the Ironical user based on their tweet contents. Due to Machine Limitations, I have used 2000 tweets from each of the Ironical and Ironical users listed in the truth value as my training set and have tested 100 users i.e, 20000 tweets to find the result for Ironical and Non-ironical users. The dataset contains:

- An XML file per author (Twitter user) with 200 tweets. The name of the XML file corresponds to the unique author id.
- A truth.txt file with the list of authors and the ground truth.

The format of the XML files is depicted in Fig. 3

```
<author lang="en">
<documents>
<document>Tweet 1 textual contents</document>
<document>Tweet 2 textual contents</document>
...
</documents>
</author>
```

Fig 3: Raw Dataset format.

6.2.2 Data Pre-processing:

We use a logistic regression with L1 regularization to reduce the dimensionality of the data. And then using Wordnet and Synsets, we find the various parts of speech for the key words and listed them.

6.2.3 Training of data:

- **We then, count the no of Synsets for a word. The word whose synsets is greater than and equal to 2, we tag the word as homonym as per process.**

- **Now we count the no of homonyms per tweets, per user and no. of homonyms used upon total no. of words used by the users for the tweets, thus we get 3 features set for our training set.**

- **Again, we find the n grams for the tweets and count the same thus got another one feature for our set.**

- **Lastly for our fifth features set we calculate the total homonyms used in the 200 tweets per users.**

Based on these features set we applied our ML model to train the data and finally obtain our training set for both the Irony and Non-Ironical users individually.

6.3 MACHINE LEARNING MODEL USED

A) **Linear Regression (LR):** The measure of the relationship between two variables is shown by the correlation coefficient. The range of the coefficient lies between -1 to +1. This coefficient shows the strength of the association of the observed data between two variables.

Linear Regression Equation is given below:

Y=a+bX

where X is the independent variable, and it is plotted along the x-axis

Y is the dependent variable, and it is plotted along the y-axis

Here, the slope of the line is b, and a is the intercept (the value of y when x = 0).

B) **Random forests** or random decision forests is an ensemble learning method for classification, regression and other tasks that operates by constructing a multitude of decision trees at training time. For classification tasks, the output of the random forest is the class selected by most trees. For regression tasks, the mean or average prediction of the individual trees is returned.[1][2] Random decision forests correct for decision trees' habit of overfitting to their training set.[3]:587–588 Random forests generally outperform decision trees, but their accuracy is lower than gradient boosted trees [citation needed]. However, data characteristics can affect their performance.[4][5].

C) **Support vector machines (SVMs)** are a set of supervised learning methods used for classification, regression and outliers detection.

The advantages of support vector machines are:

- Effective in high dimensional spaces.
- Still effective in cases where number of dimensions is greater than the number of samples.
- Versatile: different Kernel functions can be specified for the decision function. Common kernels are provided, but it is also possible to specify custom kernels.

The disadvantages of support vector machines include:

- If the number of features is much greater than the number of samples, avoid over-fitting in choosing Kernel functions and regularization term is crucial.
- SVMs do not directly provide probability estimates, these are calculated using an
- expensive five-fold cross-validation (see Scores and probabilities, below)

7 IMPLEMENTATION AND RESULT

We trained the final model using the entire dataset and used it to predict the user. We use a one-versus-restframework where a separate classifier is trained for each class and the class label with the highest predicted probability across all classifiers is assigned to each tweet. All modeling was performing using scikit-learn.

The trained set so far, we have obtained from our given data set is as below,

TABLE-4: Homonyms frequency for each user

	User	Homonym/tweet for Ironic user	Homonym/tweet forNon-Ironic user	Homonyms/Ironic user	Homonym/Non-Ironic User
0	1	87.93	47.97	14.371905	12.035119
1	2	60.65	56.78	23.143519	14.033113
2	3	85.43	56.30	23.237668	21.097353
3	4	78.63	49.68	20.251030	15.721862
4	5	89.20	57.94	21.097353	11.090000
5	6	66.40	41.32	15.721862	14.230000
6	7	83.86	53.23	22.860000	10.860000
7	8	62.12	48.89	14.320000	9.870000
8	9	87.74	49.22	16.890000	14.670000
9	10	71.48	55.68	20.230000	17.230000

TABLE-5: Frequency and Mean of homonyms for all the users

		Homonym/tweet for Ironic user	Homonym/tweet forNon-Ironic user	Homonyms/Ironic user	Homonym/Non-Ironic User
mean		77.344000	51.70100	19.212334	14.083745
std		11.229019	5.20544	3.583483	3.390317
min		60.650000	41.32000	14.320000	9.870000
25%		67.670000	48.97250	16.013896	11.326280
50%		81.245000	51.45500	20.240515	14.131556
75%		87.162500	56.14500	22.419338	15.458896

Histogram of the homonyms used by Ironic and non-ironic users with accuracy values from table is described as below,

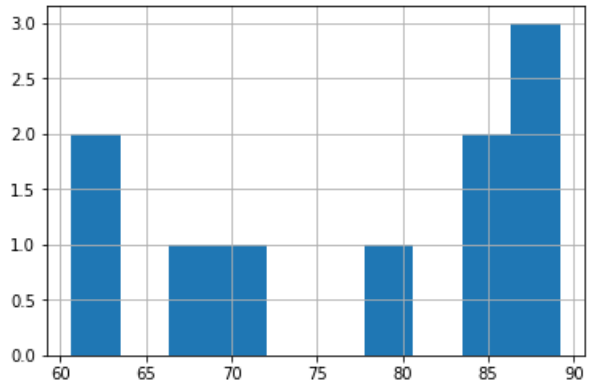

Fig-4: Homonym per tweet for Ironic

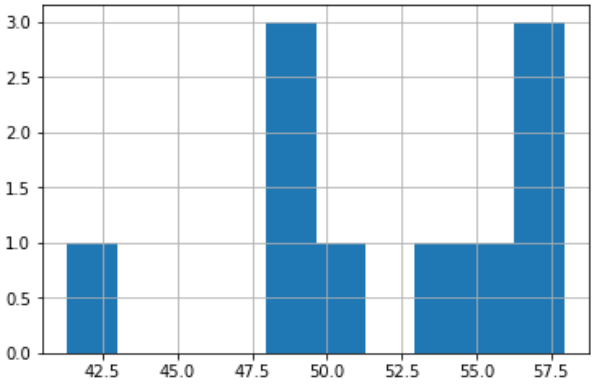

Fig-5: Homonyms/tweet for non-ironic

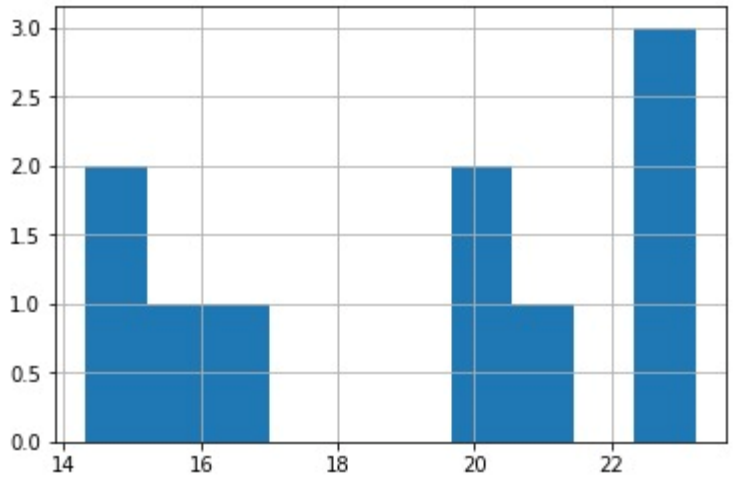

Fig-6: Homonyms frequency/user for Ironic users

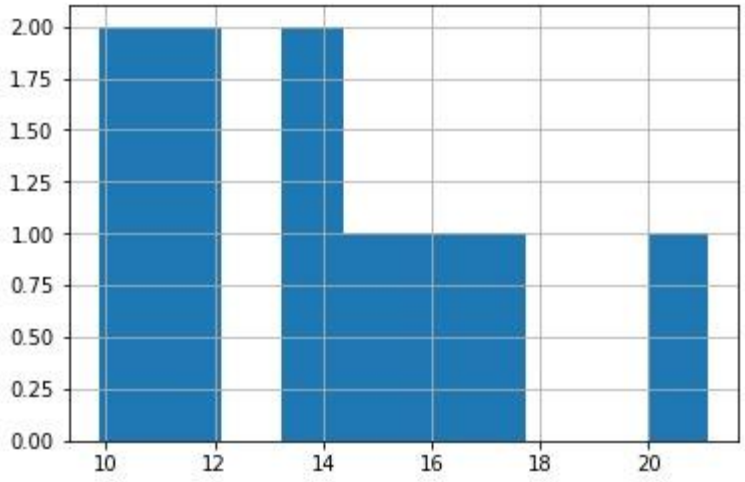

Fig-7: Homonyms Frequency/user for non-Ironic users

7.1 Implementation: Here, we have used 100 +100 users as our testing set, i.e., We have considered a total of 20000 + 20000 each for ironic and non-Ironic detection of the users.

7.1.1 On implementing the **LR model**, below result is obtained,

TABLE-6: LR MODEL PREDICTION MATRIX

Prediction Matrix	Predicted	
	Ironic	Non-Ironic
Ironic	76	24
Non-Ironic	32	68

TABLE-7: LR MODEL CONFUSION MATRX

Measure	Value	Derivations
Sensitivity	0.7600	TPR = TP / (TP + FN)
Specificity	0.6800	SPC = TN / (FP + TN)
Precision	0.7037	PPV = TP / (TP + FP)
Negative Predictive Value	0.7391	NPV = TN / (TN + FN)
False Positive Rate	0.3200	FPR = FP / (FP + TN)
False Discovery Rate	0.2963	FDR = FP / (FP + TP)
False Negative Rate	0.2400	FNR = FN / (FN + TP)
Accuracy	0.7200	ACC = (TP + TN) / (P + N)
F1 Score	0.7308	F1 = 2TP / (2TP + FP + FN)
Matthews Correlation Coefficient	0.4414	TP*TN - FP*FN / sqrt((TP+FP)*(TP+FN)*(TN+FP)*(TN+FN))

Accuracy: 72%
F1 Score: 73.08%
Precision: 70.37%

7.1.2 On implementing the **RANDOM FOREST**, below result is obtained,

TABLE-8: RANDOM FOREST PREDICTION MATRIX

Prediction Matrix	Predicted	
	Ironic	Non-Ironic
Ironic	78	22
Non-Ironic	25	75

TABLE-9: RANDOM FOREST CONFUSION MATRIX

Measure	Value
Sensitivity	0.7573
Specificity	0.7732
Precision	0.7800
Negative Predictive Value	0.7500
False Positive Rate	0.2268
False Discovery Rate	0.2200
False Negative Rate	0.2427
Accuracy	0.7650
F1 Score	0.7685

Accuracy: 76.50%
F1 Score: 76.85%
Precision: 78.00%

7.1.3 On implementing the **SVM**, below result is obtained,

TABLE-10: SVM PREDICTION MATRIX

Prediction Matrix	Predicted	
	Ironic	Non-Ironic
Ironic	84	16
Non-Ironic	17	83

TABLE-11: SVM CONFUSION MATRIX

Measure	Value
Sensitivity	0.8317
Specificity	0.8384
Precision	0.8400
Negative Predictive Value	0.8300
False Positive Rate	0.1616
False Discovery Rate	0.1600
False Negative Rate	0.1683
Accuracy	0.8350
F1 Score	0.8358

Accuracy: 83.50%
F1 Score: 83.58%
Precision: 84.00%

7.2 Accuracy comparison:

TABLE -12: Accuracy comparison

	Precision	F1 Score	Accuracy
LR	0.7037	0.7308	0.7200
RF	0.7800	0.7685	0.7650
SVM	0.8400	0.8358	0.8350

8

CONCLUSION

Irony is a difficult phenomenon to define and is not monolithic. Our classifications of Irony tend to reflect ourown subjective biases. People identify racist and homophobic slurs as hateful but tend to see sexist language as merely offensive. While our results show that people perform well at identifying some of the more egregious instances of Ironical and non-ironical, it is important that we are cognizant of the social biases that enter into our algorithms and future work aim to identify and correct these biases and also to increase our dataset quantity for a more precise result

9

FUTURE WORKS

Primarily in this project we have successfully detected the Ironical tweets and thus the user from a social analysis and thus established ourresult. On further process we are looking forward to increasing our accuracy for the same by implementing a new algorithm and to cover up the images, videos and thus providing a more accurate analysis of the hate speeches whichalso helps us to detects the spreaders as well.

Also, we are aiming on proving a live implementation of the work through an API thus to detect the same at the instance

References:

[1] Hajime Watanabe, Mondher Bouazizi, Tomoaki Ohtsuki, “Hate Speech on Twitter: A Pragmatic Approach to collectHateful and Offensive Expressions and Perform Hate Speech Detection”.

[2] Rakshita Jain, Devanshi Goel, Prashant Sahu, Abhinav Kumar, Jyoti Prakash Singh, “Profiling Hate Speech Spreaderson Twitter”.

[3] Pinkesh Badjatiya, Shashank Gupta, Manish Gupta, Vasudeva Varma, “Deep Learning for Hate Speech Detection inTweets”.

[4] Thomas Davidson, Dana Warmsley, Michael Macy, Ingmar Weber, “Automated Htae Speech Detection and theProblem of Offensive Language”.

[5] Noman Ashraf, Abid Rafiq and Sabur Butt and Hafiz Muhammad Faisal Shehzad and Grigori Sidorov and Alexander Gelbukh, “YouTube Based Religious Hate Speech and Extremism Detection Dataset with Machine Learning Baselines”.

[6] Ashish Sureka, Ponnurangam Kumaraguru, Atul Goyal, Sidhart Chhabra, “Mining YouTube to Discover ExtremistVideos, Users and Hidden Communities”.

[7] Fabio Del Vigna, Andrea Cimino, Felice Dell’Orletta, Marinella Petrocchi, and Maurizio Tesconi, “Hate me, hate menot: Hate speech detection on Facebook”.

[8] ANAT BEN-DAVID, ARIADNA MATAMOROS-FERNÁNDEZ, “Hate Speech and Covert Discrimination on Social Media: Monitoring the Facebook Pages of Extreme-Right Political Parties in Spain”.

[9] Shakeel Ahmad1, Muhammad Zubair Asghar, Fahad M. Alotaibi and Irfanullah Awan,” Detection and classifcationof social media-based extremist afliations using sentiment analysis techniques”.

[10] Njagi Dennis Gitari , Zhang Zuping1 , Hanyurwimfura Damien and Jun Long, “A Lexicon-based Approach for HateSpeech Detection”.

[11] Data set: https://pan.webis.de/clef22/pan22-web/author-profiling.html.

[12] Erik Forslid Niklas Wikén, “Automatic irony- and sarcasm detection in Social media”.

[13] Jens Lemmens and Ben Burtenshaw and Ehsan Lotfi and Ilia Markov and Walter Daelemans, “Sarcasm Detection Using an Ensemble Approach

[14] https://scikit-learn.org/stable/modules/svm.html#:~:text=Support%20vector%20machines%20(SVMs)%20are,than%20the%20number%20of%20samples.

[15] https://en.wikipedia.org/wiki/Random_forest

[16] https://www.ibm.com/in-en/topics/linear-regression#:~:text=Resources-,What%20is%20linear%20regression%3F,is%20called%20the%20independent%20variable.

[17] Accuracy, F1 Score, Precision and Recall in Machine Learning (thecleverprogrammer.com)

[18] Support-vector machine - Wikipedia

[19] Tokenize text using NLTK in python - GeeksforGeeks

[20] Sumit Gupta and Puja Halder, “A Hybrid Lexicon-Based Sentiment and Behaviour Prediction System”

.