A Project report on

GENDER IDENTIFICATION OF AUTHOR FROM TEXT USING NLP

A Dissertation submitted to JNTU Hyderabad in partial fulfillment of the academic requirements for the award of the degree.

Bachelor of Technology

In

Computer Science and Engineering

Submitted by

N. SHRAVANI
(20H51A0541)
S. CHANDANA
(20H51A0577)
B. SAI RAMAN
(20H51A0583)

Under the esteemed guidance of

Ms.M. KAMALA
(Assistant Professor)

Department of Computer Science and Engineering

CMR COLLEGE OF ENGINEERING & TECHNOLOGY

(UGC Autonomous)
*Approved by AICTE *Affiliated to JNTUH *NAAC Accredited with A^{+} Grade

KANDLAKOYA, MEDCHAL ROAD, HYDERABAD - 501401.

2020- 2024

CMR COLLEGE OF ENGINEERING & TECHNOLOGY
KANDLAKOYA, MEDCHAL ROAD, HYDERABAD – 501401

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

CERTIFICATE

This is to certify that the Major Project Phase I report entitled **"Gender Identification of Author from text using NLP"** being submitted by N.Shravani(20H51A0541),S.Chandana(20H51A0577),B.SaiRaman(20H51A0583) in partial fulfillment for the award of **Bachelor of Technology in Computer Science and Engineering** is a record of bonafide work carried out his/her under my guidance and supervision.

The results embodies in this project report have not been submitted to any other University or Institute for the award of any Degree.

Ms. M. Kamala
Assistant Professor
Dept. of CSE

Dr. Siva Skandha Sanagala
Associate Professor and HOD
Dept. of CSE

ACKNOWLEDGEMENT

With great pleasure we want to take this opportunity to express my heartfelt gratitude to all the people who helped in making this project work a grand success.

We are grateful to **Ms. M. Kamala, Assistant Professor** , Department of Computer Science and Engineering for her valuable technical suggestions and guidance during the execution of this project work.

We would like to thank **Dr. Siva Skandha Sanagala,** Head of the Department of Computer Science and Engineering, CMR College of Engineering and Technology, who is the major driving forces to complete my project work successfully.

We are very grateful to **Dr. Vijaya Kumar Koppula**, Dean-Academics, CMR College of Engineering and Technology, for his constant support and motivation in carrying out the project work successfully.

We are highly indebted to **Major Dr. V A Narayana,** Principal, CMR College of Engineering and Technology, for giving permission to carry out this project in a successful and fruitful way.

We would like to thank the **Teaching & Non- teaching** staff of Department of Computer Science and Engineering for their co-operation

We express our sincere thanks to **Shri. Ch. Gopal Reddy**, Secretary, CMR Group of Institutions, for his continuous care.

Finally, We extend thanks to our parents who stood behind us at different stages of this Project. We sincerely acknowledge and thank all those who gave support directly and indirectly in completion of this project work.

N. Shravani	20H51A0541
S. Chandana	20H51A0577
B. Sai Raman	20H51A0583

TABLE OF CONTENTS

List of Figures

List of Tables

ABSTRACT

The emergence of Natural Language Processing (NLP) techniques has revolutionized the field of text analysis, enabling researchers to delve deeper into understanding human communication. This project aims to contribute to the field by developing an advanced model for the automatic identification of the gender of authors based on their written text. Gender identification from text is a complex task as it involves the extraction of subtle linguistic patterns and nuances inherent to different genders.

This research project leverages a diverse array of NLP techniques, including text preprocessing, feature extraction, and various machine learning algorithms to train and validate a robust gender identification model. By utilizing a large dataset of text samples authored by individuals from different genders, the model is trained to recognize and decipher the underlying linguistic characteristics specific to male and female writers. Additionally, special attention is given to the nuanced challenges associated with identifying gender in non-binary and transgender individuals.

The proposed model not only demonstrates its effectiveness in accurately predicting the gender of the author based on text data but also provides insights into the linguistic differences between genders, shedding light on the broader societal implications of language use. Furthermore, the project explores the ethical considerations related to the potential implications of automated gender identification and emphasizes the significance of ensuring fairness, transparency, and privacy in the deployment of such models.

Ultimately, this project endeavors to contribute to the advancement of NLP research, promoting a deeper understanding of gender-specific linguistic patterns and facilitating the development of more inclusive and equitable NLP applications.

CHAPTER 1
INTRODUCTION

CHAPTER 1

INTRODUCTION

1.1. Problem Statement

The objective of this research project is to develop a application despite the increasing prevalence of Natural Language Processing (NLP) applications, the accurate identification of an author's gender solely based on their written text remains a challenging and nuanced task. Existing research often overlooks the subtleties and complexities associated with gender identification, leading to a lack of robust models that can effectively discern gender-specific linguistic patterns from text data. Additionally, the presence of non-binary and transgender individuals further complicates the accurate classification of gender, highlighting the need for a more inclusive approach in gender identification using NLP.

1.2 Research Objective

- Develop a robust NLP model capable of accurately identifying the gender of authors based on textual data, considering a diverse set of linguistic features that capture the subtle nuances in language use specific to different genders.

- Investigate and incorporate non-binary and transgender linguistic patterns into the gender identification model to ensure inclusivity and accuracy in the classification of authors with diverse gender identities.

- Explore the ethical implications associated with automated gender identification, emphasizing the importance of ensuring fairness, transparency, and privacy protection in the development and deployment of the NLP model.

- Analyze the linguistic differences between genders in written text, aiming to provide insights into the broader societal implications of language use and contribute to a deeper understanding of gender-specific communication patterns.

- Evaluate the performance of the developed model using a comprehensive dataset, comparing its accuracy, precision, and recall with existing gender identification methods to demonstrate its effectiveness and potential for practical applications in various NLP domains.

- Propose guidelines and best practices for the ethical deployment of gender identification models, highlighting the importance of mitigating biases, promoting inclusivity, and safeguarding individual privacy in NLP-based gender analysis.

CHAPTER 2
BACKGROUND WORK

CHAPTER 2

BACKGROUND WORK

2.1 Gender Classification using Twitter Text Data

2.1.1 Introduction

Twitter is one of the most famous social networking websites which allows online users to read and post 280- character messages on their website. These messages are generally referred to as tweets. The message size restriction is one of the reasons Twitter has become very popular among social media users. Twitter provides an equal opportunity to global online users to connect with other users including top celebrities, politicians, famous people, etc. and follow them regularly via reading their tweets. The increasing popularity of social media has created a unique opportunity to learn about human society on a large scale.

According to Omnicore agency, which is one of the well-known marketing websites, monthly active twitter use reached more than 330 million users, users tweet more than 500 million tweets per day and about 139 million people actively use Twitter on a daily basis. It has also been discovered that approximately 34% of Twitter users are females and more than 66% are male. Twitter has a default option to display the tweets publicly, however, users can put restrictions on messages so they can only be viewed by approved followers. Twitter's main ideology is focused on sharing tweets publicly. According to Kvamme, over 90% of Twitter users prefer to make their interactions publicly accessible.

With increasing use of shared content on social media, classifying gender is gaining significant interest from marketing companies and government. Extracting knowledge from this hidden content could be very beneficial to various Gender classification can be considered a problem of binary classification, for instance, the goal is to classify the object into two groups based on some features. Public availability of such user text data provides the researcher with ample opportunity to find the patterns within this data. Such unstructured/structured text data can be normalised using different text pre-processing and Natural language processing (NLP) techniques. NLP, which is a subfield of Artificial Intelligence (AI), provide machines the ability to understand spoken/written human language. NLP includes different ways to analyse and understand this information and obtain meaning from human language automatically.

2.1.2 Merits,Demerits and Challenges

Advantages

- Large Data Availability.
- Real-Time Data
- Linguistic Informality.
- Diversity in Topics.

Disadvantages

- Noisy Data.
- Limited Context.
- Limited Text Length.
- Privacy Concerns.
- Biases and Stereotypes
- Validation and Accuracy

2.1.3 Implementation of Gender Classification using Twitter Text Data

Data Pre-Processing

The data pre-processing is a very critical step because better results can only be achieved with good quality of data. This is achieved by following steps such as feature engineering, visualisations of data and ML classifiers. Since the social media data is unstructured or in other words it's raw and very noisy it requires robust cleaning. The main objective of this step is to remove noisy and inconsistent tweet data. Tweets that carry very little weighting in text context, for example numbers, special character, punctuations, hashtags, emojis, smileys, extra blank space, etc. need to be removed. This step also involves filtering the tweet length required for feature engineering. Once the raw data is filtered and cleaned then we need to ensure tweets are written in the English language, therefore further pre-processing was completed. Finally removing stop words by using NLTK library during run time.

Data Mining

A Once data transformation has completed with multidimension features, the next step is to classify the updated data on ML models. A supervised ML approach is considered in this research to predict the output label class correctly as per previous research. The following classifiers were considered in this research; LR, MLP, SVM, Naïve Bayes, RF and XGBoost. These are simple yet effective algorithms for binary classification. A base-line is created using a simple pipeline consisting of the TF-IDF and LR model. Further various algorithms are combined with different feature engineering techniques and were tested to compare the results with the baseline.

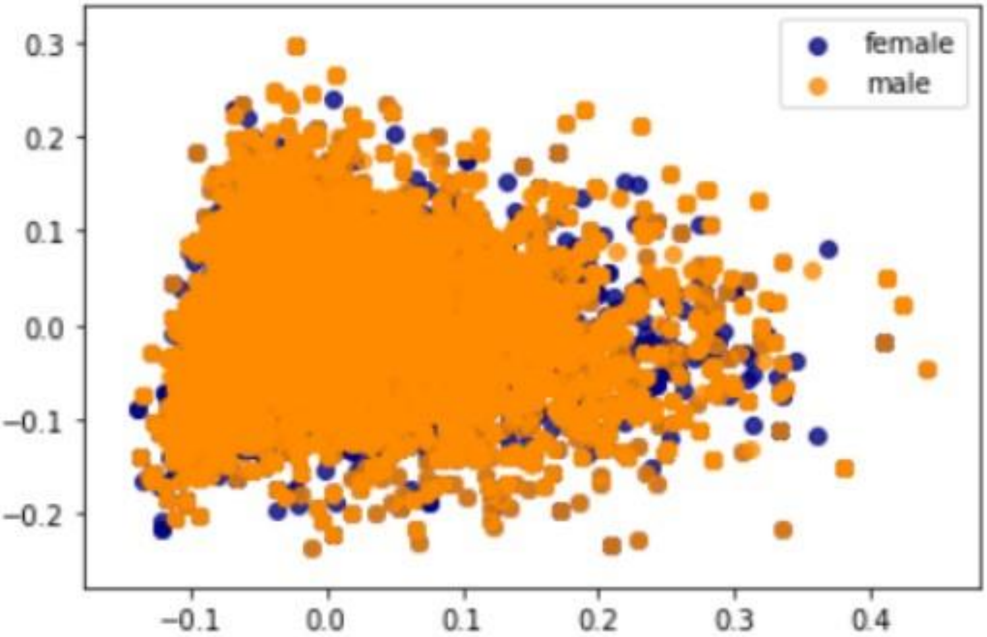


Fig:2.1.3.1 Male and Female Tweets Plot using PCA

Utilising Machine Learning Models

This research utilised six different ML classification algorithms to predict the gender class on a featured Twitter dataset. SVM, Naïve Bayes, LR have proven to be very successful and widely used in the past with binary classification problems. These techniques are also well-known for good predictive performance regardless the amount of data. Naïve Bayes is known for its simplicity and ease of implementation. This technique has also outperformed alternatives in real applications when it comes to reliability, easy and efficient classification. Similarly, XG boost has the best model performance and execution speed. Additional techniques such as RF and MLP are also tested for this research due to comparison purposes to analyse different result achieved when compared with other well-known classifiers

Dimensionality Reduction Once the pre-processed step is completed, there are approx. 21,021 tweets (male &female). To plot the data dimensionality reduction techniques were used. A number of features

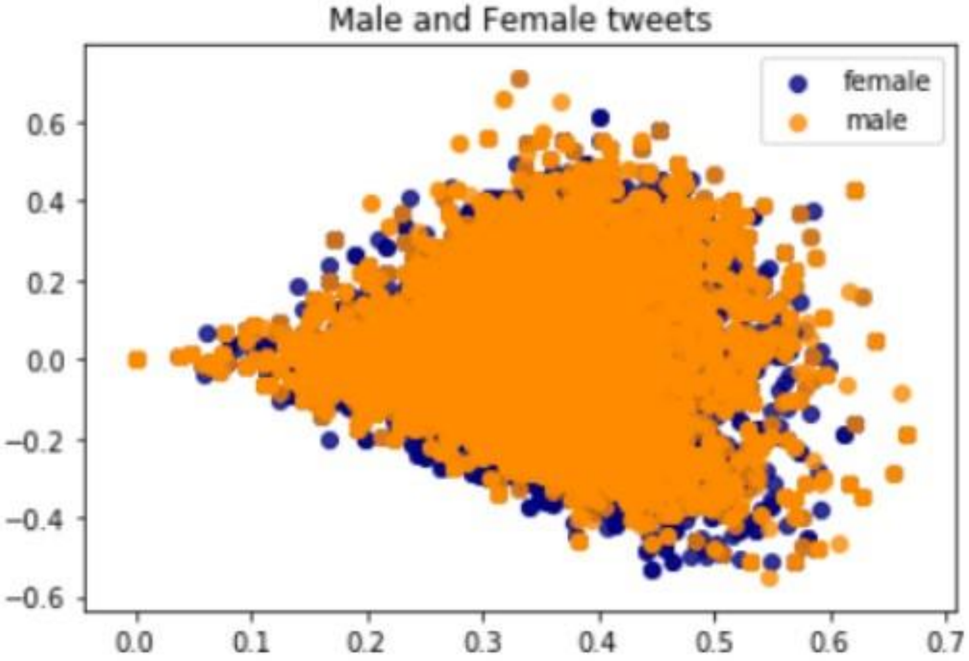
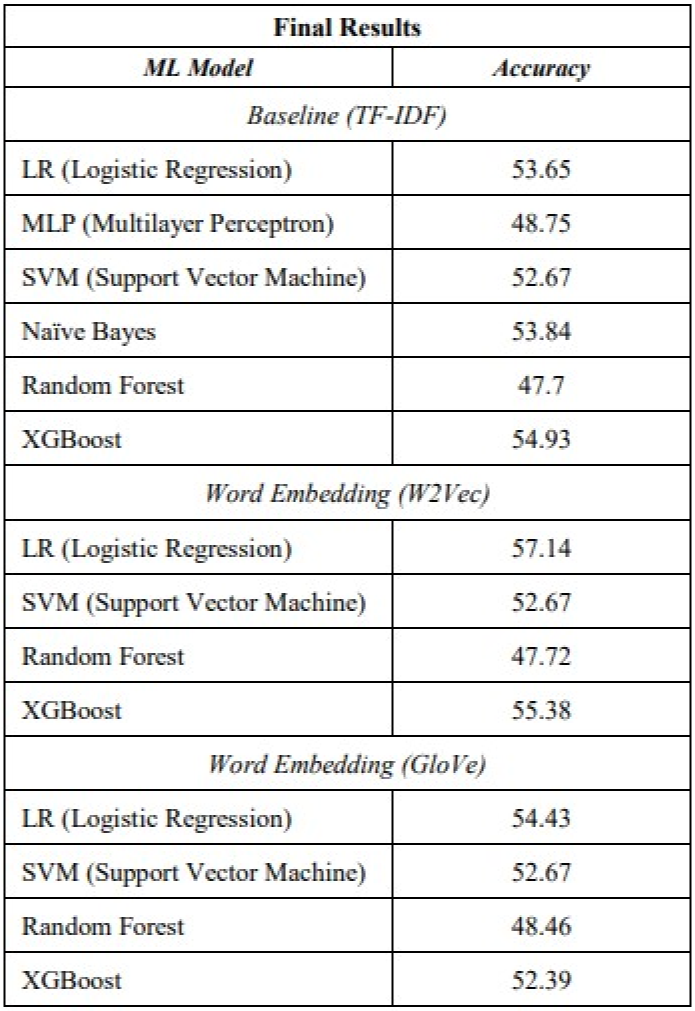


Fig:2.1.3.2 Male and Female Tweets Plot using truncated SVD

were set to two so that it can be plotted on the X-Y axis. The purpose of plotting data is to get an insight around how separable the two classes are. If the two classes are separable via a linear or no-linear boundary, then it should be easy to train a model to achieve desirable performance. Conversely, if the classes are not easily separable, then it could be difficult to achieve desirable performance using simple techniques.

The graph shows the data for both genders is overlapping therefore, it’s difficult to find if some unique words are used by males and females individually. In other words, there is not much variance in the feature data among male and female tweets. The Truncated SVD (Singular Value Decomposition) technique computes the linear dimensionality through means of truncated SVD like PCA but works on sample vectors directly in place of finding covariance matrices. The feature transformed data was then plotted where X and Y represents Male and Female tweets. The graph 2 is very similar to PCA due to both maintaining the largest amount of variance.

A. TF-IDF Once the data is pre-processed then the next step will be to apply the feature extraction technique on to the updated dataset. Feature extraction techniques help the ML algorithms to understand the text by converting it into numbers. TF-IDF is known as one of the most efficient statistical weighted measured approaches used for feature extraction and is provided in the Scikit-learn library. B. Word Embedding The word embedding (W2Vec. GloVe) techniques are expected to detect extra semantic features and reduce dimensions in detecting the gender class. Previous research is mostly involved without traditional BOW approaches e.g. BOW, TF-IDF, N-grams, etc. Therefore, the approach of this research is representing text through the VSM which is more semantically rich .



Final Results	
ML Model	***Accuracy***
Baseline (TF-IDF)	
LR (Logistic Regression)	53.65
MLP (Multilayer Perceptron)	48.75
SVM (Support Vector Machine)	52.67
Naïve Bayes	53.84
Random Forest	47.7
XGBoost	54.93
Word Embedding (W2Vec)	
LR (Logistic Regression)	57.14
SVM (Support Vector Machine)	52.67
Random Forest	47.72
XGBoost	55.38
Word Embedding (GloVe)	
LR (Logistic Regression)	54.43
SVM (Support Vector Machine)	52.67
Random Forest	48.46
XGBoost	52.39

Fig:2.1.3.3 COMPARING MULTIPLE ML TECHNIQUES.

W2Vec and GloVe models were used as part of the genism library. This research will use a pre-trained vector as these models are stronger regarding word semantics and are trained on a large corpus such as Wikipedia. W2Vec group the vector of similar words in vector space for every unique word which helps in detecting mathematical similarity of word features without any human intervention. Fortunately, genism provides a pre-trained vector on the Google News dataset for W2Vec. The model contains 300-dimensional vectors for 3 million words and phrases. Similarly, genism provides GloVe's pre-trained model on 2 billion tweets, 27 billion tokens and 1.2 million pieces of vocabulary. For this experiment, the mean of all word vectors in the tweet has been taken since advanced Deep Learning models were not applied. The length of the resultant vector isthe same for W2vec (Google) 300 and GloVe (Twitter) model 200. These vectors are very good at capturing the semantics and even analogies between different words. The same process is repeated to obtain the vector representation for all tweets data.

2.2 The Gender Gap Tracker: Using Natural Language Processing to measure gender bias in media

2.2.1 Introduction

In contemporary society, understanding the subtle nuances and pervasive effects of gender bias in media has become a critical endeavor. The perpetuation of gender stereotypes and inequalities through various forms of media content has garnered widespread attention, emphasizing the urgent need for robust tools to measure and track gender disparities. To address this imperative, the Gender Gap Tracker harnesses the power of Natural Language Processing (NLP) to provide a comprehensive and dynamic analysis of gender bias in media content.

The Gender Gap Tracker represents a groundbreaking initiative aimed at scrutinizing textual data from diverse media sources, including news articles, online publications, and social media platforms. Leveraging advanced NLP techniques, this project seeks to quantify and evaluate the linguistic manifestations of gender bias, enabling a nuanced understanding of how language usage influences societal perceptions and attitudes toward gender roles and identities.

With an overarching commitment to fostering a more equitable and inclusive media landscape, the Gender Gap Tracker endeavors to achieve the following key objectives: to systematically identify gender-specific language patterns, to quantify the prevalence of gender bias across different media genres, and to provide actionable insights that can inform stakeholders in the media industry, policy makers, and the public at large.

This project introduction lays the foundation for a comprehensive exploration of the Gender Gap Tracker, highlighting its significance in promoting a more conscientious and informed approach to content creation, consumption, and regulation in the media sphere. By employing cutting-edge NLP methodologies, this initiative aspires to contribute to the advancement of a more equitable and unbiased media landscape, thereby fostering a more inclusive and progressive society.

2.2.2 Merits,Demerits and Challenges

Advantages

- Comprehensive Analysis.
- Objective Measurement
- Timely Insights

- Quantitative Assessment.

Disadvantages

- Contextual Complexity.
- Algorithmic Limitations.
- Data Accessibility.
- Subjectivity in Interpretation.
- Ethical Considerations.

2.2.3 Implementation on The Gender Gap Tracker: Using Natural Language Processing to measure gender bias in media

- Data Collection and Curation: Gather a diverse dataset of media content, including news articles, online publications, and social media posts, ensuring representation across different genres, sources, and time periods. Curate the dataset to ensure a balanced representation of content across various demographic and thematic categories.

- Preprocessing and Text Cleaning: Preprocess the collected textual data to remove noise, such as special characters, punctuation, and irrelevant formatting. Perform text cleaning tasks, including tokenization, lemmatization, and stemming to standardize the text data for further analysis.

- Gender Bias Indicator Development: Develop a comprehensive set of linguistic and contextual indicators to detect and measure gender bias within the textual data. This may involve the creation of specific linguistic metrics and feature sets that capture gender-specific language usage and stereotypical representations.

- NLP Analysis and Algorithm Design: Implement advanced NLP algorithms, such as sentiment analysis, topic modeling, and word embedding techniques, to identify and quantify instances of gender bias within the media content. Design and develop custom algorithms tailored to the specific context of the project and the defined indicators of gender bias.

- Bias Quantification and Visualization: Quantify and visualize the identified gender bias patterns and trends within the media content using descriptive statistics, data visualization techniques, and interactive dashboards. Generate visual representations, such as heatmaps, graphs, and trend charts, to facilitate a comprehensive understanding of the prevalent gender biases in different forms of media.

Interpretation and Analysis: Interpret the results of the NLP analysis to understand the underlying causes and implications of gender bias in the media. Conduct in-depth textual analysis and contextual interpretation to identify the socio-cultural and linguistic factors contributing to the perpetuation of gender stereotypes and inequalities.

- Reporting and Communication: Prepare detailed reports and presentations outlining the findings, insights, and recommendations derived from the Gender Gap Tracker analysis. Communicate the results to stakeholders, including media organizations, policymakers, and advocacy groups, to raise awareness and promote informed discussions about gender bias in media content.
- Iterative Refinement: Continuously refine the NLP algorithms and bias detection mechanisms based on feedback, new data, and emerging trends in media content. Implement iterative improvements to enhance the accuracy, reliability, and scalability of the Gender Gap Tracker's analytical capabilities

. The Tracker monitors mainstream Canadian media, seven English-language news sites (a French Tracker is in development), motivating them to improve the current disparity. By openly displaying ratios and raw numbers for each outlet, we can monitor the progress of each news organization towards gender parity in their sources. Fig 1 shows a screenshot of the live page. In addition to the bar charts for each organization and the doughnut chart for aggregate values, the web page also displays a line graph, charting change over time (see Fig 2 below).

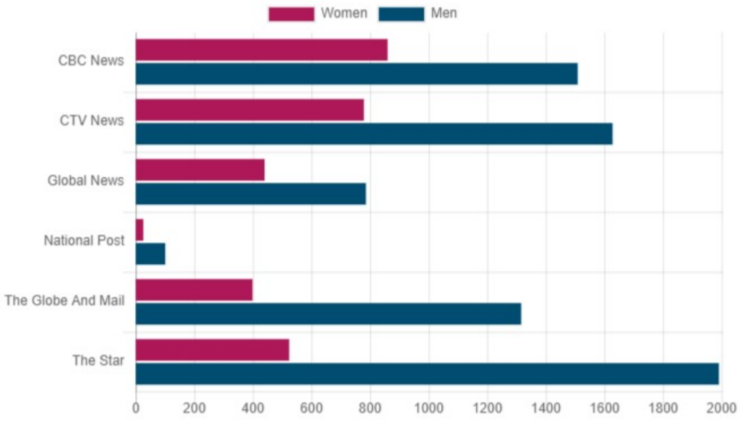


Fig 2.2.3.1:H Ratio of men and women sources by news outlet.

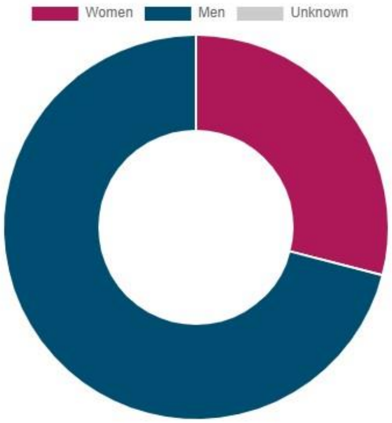


Fig 2.2.3.1 : Aggregate ratio of men and women sources

Counts and percentages of male vs. female sources of opinion across seven news outlets.
Dates: October 1, 2018 to September 30, 2020. Female sources constitute less than 30% of the sources overall. CBC News (blue line) and HuffPost Canada (green line) show a better gender balance compared to other outlets; The Globe and Mail (light blue) and The National Post (orange) are at the bottom, quoting women less than 25% of the time.

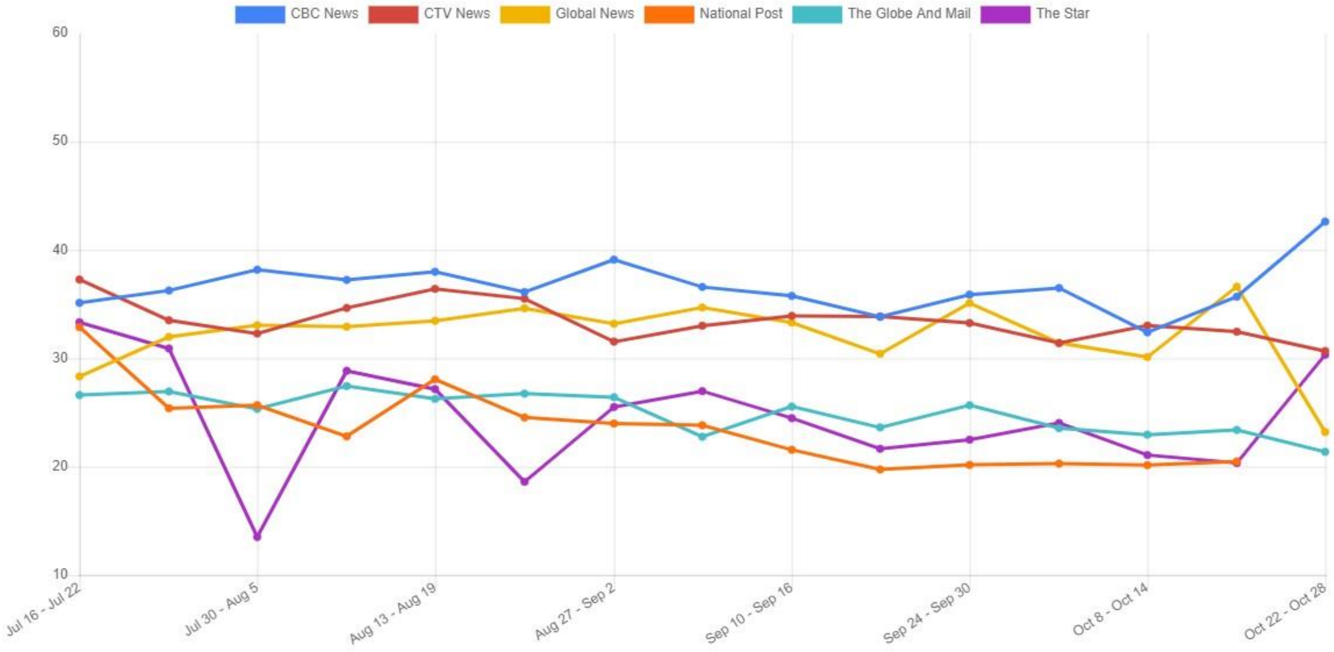


Fig 2.2.3.3:Percentage of Women Quoted in Canadian Online News Media

In this paper, we describe the data collection and analysis process, provide evaluation results and a summary of our analysis and observations from the data. We also outline other potential uses of the tool, from quantifying gender representation by news topic to uncovering emerging news topics and their protagonists. We start, in Related work, with a review of existing literature on quotation patterns, extracting information from parsed text, and potential biases in assigning gender to named entities. More detail for each of those steps is provided in the S1 Appendix. Throughout the development of the Gender Gap Tracker, we were mindful of the need for accuracy, in both precision and recall of quotes, but also in terms of any potential bias towards one of the genders. In order to ensure that the Gender Gap Tracker provides as accurate a picture as possible, we have performed continuous evaluations. We describe that process in the section on Evaluation. The section Analysis and observations answers the most important questions that we posed at the beginning of the project: Who is quoted, in what pro-portions? We add more nuanced analyses about the relationship between author gender and the gender breakdown of the people those authors quote. Finally, Conclusion offers some reflections on the use of the Gender Gap Tracker as a tool for change, also discussing future improvements and feature additions.

2.3 Gender Bias in Multilingual Embeddings and Cross-Lingual Transfer

2.3.1 Introduction

In recent years, the development and utilization of multilingual embeddings in natural language processing (NLP) have gained significant attention. Multilingual embeddings enable the representation of words or sentences from different languages in a shared semantic space, facilitating cross-lingual transfer learning and improving the performance of various NLP tasks across multiple languages. However, the emergence of gender bias within these multilingual embeddings poses a substantial challenge to the equitable and unbiased treatment of various genders across different languages.

Gender bias in multilingual embeddings can manifest in several ways. Firstly, the transfer of biases from high-resource languages to low-resource languages can perpetuate and amplify existing biases, thereby hindering the development of fair and inclusive NLP systems. Secondly, the inadequate representation of gender-specific nuances and cultural variations in multilingual embeddings can lead to the reinforcement of stereotypes and discriminatory practices. Consequently, these biases can have far-reaching implications, including exacerbating gender disparities, reinforcing societal prejudices, and perpetuating discriminatory outcomes in various downstream applications such as machine translation, sentiment analysis, and information retrieval.

This research aims to comprehensively investigate the gender bias present in multilingual embeddings and its implications for cross-lingual transfer learning. By analyzing the underlying mechanisms that contribute to the propagation of gender bias across different languages, we seek to develop effective mitigation strategies and techniques to promote fair and unbiased NLP models. Furthermore, this study emphasizes the significance of developing culturally sensitive and linguistically diverse datasets to ensure the equitable representation of all genders and linguistic communities in multilingual NLP applications.

Through an in-depth exploration of the challenges posed by gender bias in multilingual embeddings, this research endeavors to foster the development of more inclusive, equitable, and culturally aware NLP systems. By highlighting the critical role of ethical considerations and responsible AI practices in mitigating gender bias, this study contributes to the ongoing efforts to build robust and unbiased NLP frameworks that uphold principles of fairness, inclusivity, and social justice across diverse linguistic and cultural contexts.

2.3.2 Merits,Demerits and Challenges

Advantages

- Cross-Lingual Transfer.
- Resource Efficiency.
- Cultural Understanding.

Disadvantages

- Biased Representation.
- Cultural Homogenization.
- Data Imbalance.
- Data Collection and Annotation

2.3.3 Implementation on Gender Bias in Multilingual Embeddings and Cross-Lingual Transfer

The Implementation of strategies to address gender bias in multilingual embeddings and cross-lingual transfer involves a multifaceted approach that integrates ethical considerations, linguistic expertise, and technical advancements. The following are key steps for implementing these strategies:

Data Collection and Curation:

Collect and curate a diverse and balanced dataset that represents various genders and linguistic communities.

Ensure the inclusion of gender-balanced data points across different languages to mitigate data bias.

Preprocessing and Embedding Techniques:

Apply preprocessing techniques to identify and remove biased language patterns and stereotypes.

Utilize advanced embedding techniques that incorporate gender-neutral representations to minimize gender bias in the embedding space.

Bias Detection and Mitigation:

Employ bias detection algorithms to identify and measure gender bias within multilingual embeddings.

Implement debiasing techniques, such as neutralizing or equalizing gender-specific embeddings, to reduce biased associations and stereotypes in the embedding space.

Language-Specific Nuances and Contextual Understanding:

Incorporate linguistic expertise and cultural knowledge to account for gender-specific nuances and socio-cultural contexts within different languages.

Develop language-specific models that capture the intricacies of gender dynamics and cultural variations, promoting

more accurate and inclusive representations.

Evaluation and Validation:

Establish comprehensive evaluation metrics and benchmarks to assess the effectiveness of bias mitigation techniques in promoting fair and inclusive cross-lingual transfer learning.

Conduct rigorous validation experiments to measure the performance and generalizability of debiased multilingual embeddings across various NLP tasks and languages.

Community Engagement and Ethical Considerations:

Engage with diverse linguistic communities and stakeholders to understand their concerns and perspectives on gender representation in multilingual NLP applications.

Uphold ethical standards and guidelines to ensure the respectful and inclusive treatment of all genders and cultural identities within the development and deployment of multilingual embeddings and cross-lingual transfer learning models.

Continuous Monitoring and Improvement:

Establish mechanisms for continuous monitoring of gender bias in multilingual embeddings and cross-lingual transfer learning models.

Regularly update and refine the debiasing techniques and algorithms to adapt to evolving linguistic, cultural, and societal dynamics.

By implementing these comprehensive strategies, researchers and practitioners can foster the development of more equitable and unbiased multilingual NLP systems that promote cultural inclusivity, linguistic diversity, and fair representation of all genders across diverse language communities.

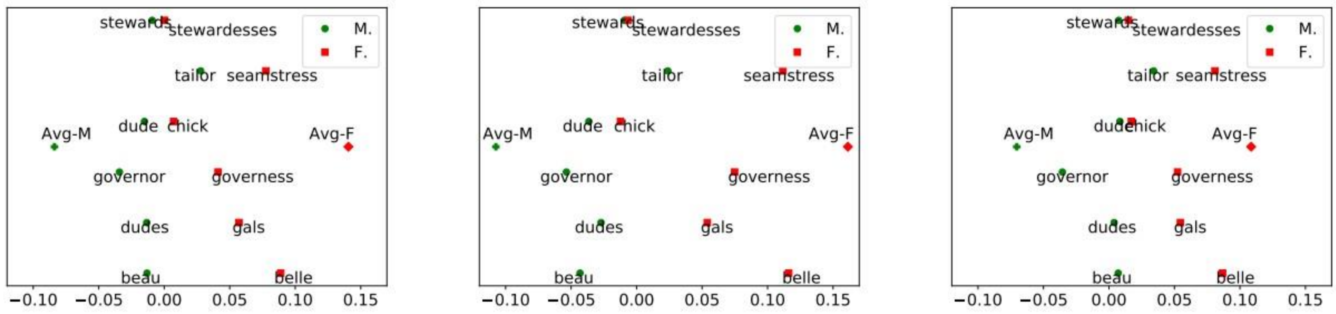
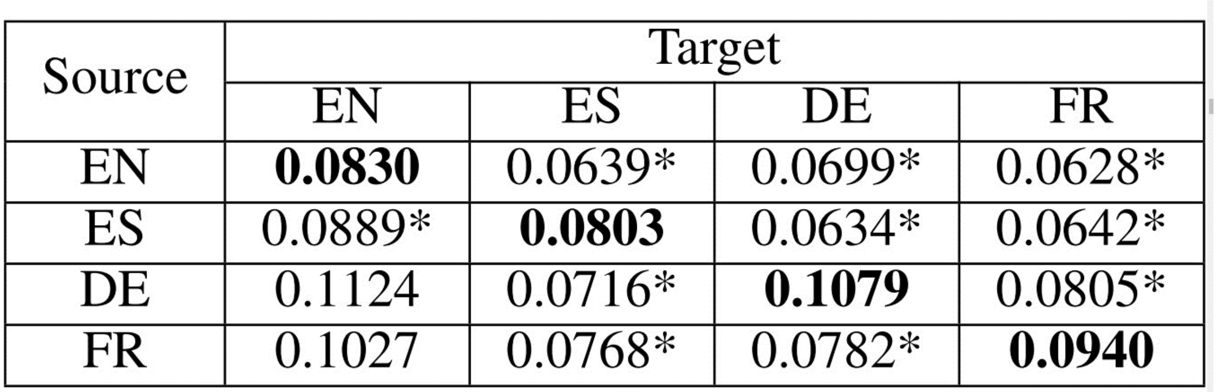


Fig 2.3.3.1:(a) Original es embeddings. (b) In es-en embeddings (c) In es-de embeddings.

As mentioned in Sec. 1, multilingual word embeddings can be generated by first training word embeddings for different languages individually and then aligning those embeddings to the same space. During the alignment, one language is chosen as target and the embeddings from other languages are projected onto this target space. We conduct comprehensive analyses on the MIBs dataset to understand: 1) how gender bias exhibits in embeddings of different languages; 2) how the alignment target affects the gender bias in the embedding space; and 3) how the quality of multilingual embeddings is affected by choice of the target language. For the monolingual embeddings of individual languages and the multilingual embeddings that used English as the target language

For all the other languages, we get the corresponding masculine and feminine terms by using online translation systems, such as Google Translate. We refer to the words that have both masculine and feminine formats in EN



Source	Target			
	EN	ES	DE	FR
EN	**0.0830**	0.0639*	0.0699*	0.0628*
ES	0.0889*	**0.0803**	0.0634*	0.0642*
DE	0.1124	0.0716*	**0.1079**	0.0805*
FR	0.1027	0.0768*	0.0782*	**0.0940**

Fig 2.3.3.2: inBias score before and after alignment to different target spaces.

CHAPTER 3

RESULTS AND DISCUSSION

CHAPTER 3

RESULTS AND DISCUSSION

PERFORMANCE MATRICS

3.1 Initially, the main dataset is split with train and test sets in an 80:20 ratio. The training set contains 16,816 tweets data while the testing set has 4,205 tweets. Training data is used to fit the model to it and then evaluated with totally unseen data to get the evaluation metrics as little bias as possible. Afterwards, the baseline TF-IDF feature extraction technique is applied to convert the text data into numeric data then fed to the ML classifier to predict the gender class. Finally, word embedding features are utilised to train ML models with more semantic rich representations. Since Naïve Bayes and MLP models do not accept the non-negative values created in word embedding pre-trained models, it is not possible to use these models, and hence were removed from the results. Table I shows the summary of all the experiments including baseline and Word embeddings (W2Vec, GloVe). The table shows good improvement in the accuracy score once the word embedding model has been introduced with the ML classifiers. The LR model achieved the highest accuracy score of 57.14% whereas the baseline approach only had 53.65% accuracy, an improvement of 6.5% increase.

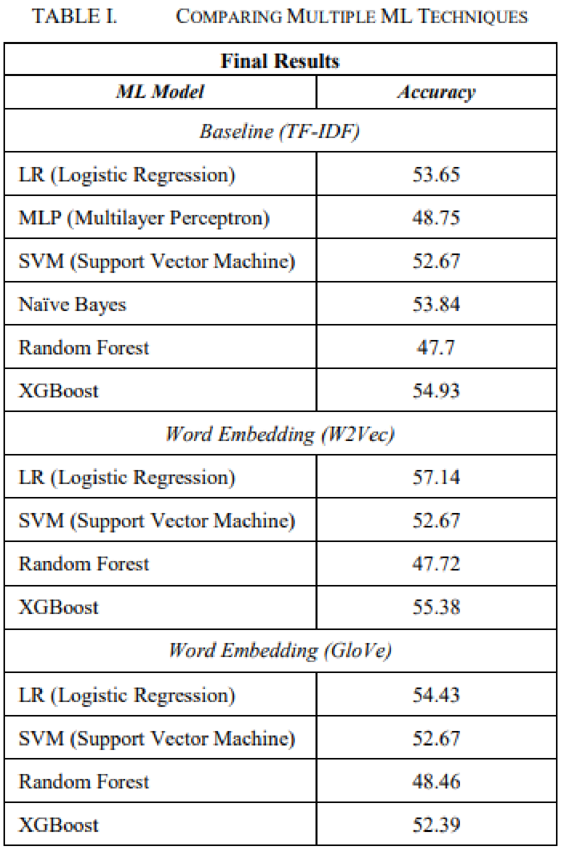


TABLE I. COMPARING MULTIPLE ML TECHNIQUES

Final Results	
ML Model	***Accuracy***
Baseline (TF-IDF)	
LR (Logistic Regression)	53.65
MLP (Multilayer Perceptron)	48.75
SVM (Support Vector Machine)	52.67
Naïve Bayes	53.84
Random Forest	47.7
XGBoost	54.93
Word Embedding (W2Vec)	
LR (Logistic Regression)	57.14
SVM (Support Vector Machine)	52.67
Random Forest	47.72
XGBoost	55.38
Word Embedding (GloVe)	
LR (Logistic Regression)	54.43
SVM (Support Vector Machine)	52.67
Random Forest	48.46
XGBoost	52.39

Fig:3.1.1

In the case of word embedding models, the performance metrics for the majority of models are improved, the LR model achieved much better results with W2Vec vectors. It also proves that word embedding techniques can detect the semantic meaning, an analogy for different words that possibly helped in improving the performance metrics. Finally, since our dataset is a unique, random, unbiased, diverse and good representation of social media, this study did not get the same results as compared to the results achieved in similar research completed in past. The dataset curated in this research is truly random and unbiased in a way that it is not curated from any specific groups like celebrities, politicians, etc. The datasets used contain millions of data points that should impact the accuracy of trained models. Regarding getting high accuracy with fewer tweets, it was observed that Miller's [25] research with gender prediction involved manually labelling data as male or female tweets.

3.2. In this section, we provide statistics on the data extracted from the seven news outlets, processed and tagged by the Gender Gap Tracker in the time frame of October 1, 2018 to September 30, 2020, 24 months of data and about 613,000 news articles. All numbers are based on the calculations of the Gender Gap Tracker version 5.3 (the most recently released version at the time of publication of this

Male vs. female sources

Shows the statistics available on the Gender Gap Tracker dashboard online. The aggregated counts and ratios of female vs. male sources across different news outlets within the time interval of October 2018 to September 2020 are presented in the bar and the doughnut charts at the top. The bottom line graph shows the percentage of women quoted in the publications of each outlet week by week. Most numbers are in the range of 20 to 30 percent, meaning that women are consistently quoted far less often than men. While some outlets such as *Huffington Post* and *CBC News* are more gender-balanced than others, such as *The National Post* and *The Globe and Mail*, the numbers suggest that, overall, media outlets disproportionately feature male voices. This may be the result of unconscious bias on the part of the reporters (e.g., reaching out to men more often than to women, when a choice exists). We, of course, also know it is a result of societal bias. In a context where 71% of the Members of Parliament are male , it is natural to expect that we hear more often from male politicians. The fact that the current (in 2020) federal cabinet is gender-balanced probably helps. It does not, however, make up for the fact that the person at the top is a man. As shown in Table 4, Justin Trudeau, the Prime Minister, is quoted 8.3 times more often than Chrystia Freeland, arguably the most prominent woman politician in the country. At the top of the list of women is Bonnie Henry, the Public Health Officer for the province of British Columbia, a reflection of how important public health officers have become in the COVID-19 pandemic. And, clearly, Donald Trump is the most quoted person by far in that time period. Perhaps the *style* of a person's statements, in addition to their content, makes the press more likely to find them quotable.

3.3. Although IAT measurements report GenS, which measures the association between men with sciences and women with humanities as a metric to measure social gender biases in human cognition, historically men have dominated the humanities in addition to science and engineering fields. Most notable philosophers, historians, linguists, writers, poets and artists throughout the history have been overwhelmingly men while women have always been more associated with domestic roles. Moreover, Caliskan et al. [10] provide evidence for a masculine default in English word embeddings, where various semantic domains are strongly associated with men. Therefore, it should not come as a surprise that GenS effect sizes are consistently smaller than GenC measurements. Furthermore, if the distribution of grammatical gender in a language is not balanced, the magnitude of grammatically feminine and masculine signals might be different. Accordingly, using the masculine and feminine version of a word might not normalize the grammatical gender signals. In our case, the grammatical gender subspace resulting from SVC may not contain feminine and masculine grammatical gender information at an equal rate. This may explain why after disentanglement, inanimate nouns still deviate from grammatical gender neutrality as shown in Figure 7. Overcoming this imbalance requires evaluating grammatical gender signals while taking other parameters such as word frequency and grammatical gender distribution in a language into account.

CHAPTER 4
CONCLUSION

CHAPTER 4

CONCLUSION

This research focused on finding a better system that could provide an autonomous solution of detecting gender class (male or female) by using only the tweet's text data. Initially, the study explores how text data is increasing day by day and the role of social media in generating billions of terabytes every single day. Twitter is one of the most popular social media platforms that generate a lot of text data and received a lot of research attention with such data as a means of understanding, predicting real-life phenomena, monitoring, etc. Also, it discussed the importance of such data for several organisations.

Moreover, a brief discussion of the research problem is deliberated rearding how text data can be challenging to classify if it's from male or female when it comes from different domains and unstructured format. The review of related literature including traditional BOW algorithms such as N-grams, TF, TF-IDF, etc. are discussed along with techniques that can help in solving an NLP problem with text classification. Recent research on gender classification using different techniques such as Word ngram, Naïve Bayes, LR, SVM, Perceptron are discussed as well as the impact these studies have on NLP. Furthermore, the needs and importance of NLP for detecting gender classification are discussed to meet the research objective. Dimension reduction techniques (SVD, PCA) also helped to understand the data used to predict the gender class. Finally, the challenges are drawn for the associated main applications. The importance of word embedding is also noticed and realised for feature extractions which can help in extracting semantic meaning from text data and improving model accuracy.

Based on the advantages and disadvantages of recent research, the predictive ML models (LR, MLP, SVM, Naïve Bayes, RF, and XGBoost) are tested on the test set with a model trained on training data. These models are tested with simple TF-IDF and word embedding (W2Vec, GloVe) features to detect the gender classification from tweet text. It is a different approach as compared to the previous research which is more focused on traditional feature extraction techniques such as BOW, TF-IDF, etc. The traditional BOW technique is only focused on counting the word frequency in a document. The best predictive model (W2Vec + LR) helped in optimising the accuracy matrices where the baseline model (simple TF-IDF + LR) is outperformed. Based on the results of this research, it is proven that classifying gender is more predictable with advanced feature extraction techniques rather than using traditional BOW approaches that do not take semantic meaning, analogy and context around the nearby words into consideration.

REFERENCES

REFERENCES

[1] K. Gligorić, A. Anderson, and R. West, 'How Constraints Affect Content: The Case of Twitter's Switch from 140 to 280 Characters', ArXiv180402318 Cs, Apr. 2018.

[2] J. A. Lopes Filho, R. Pasti, and L. De Castro, 'Gender Classification of Twitter Data Based on Textual Meta-Attributes Extraction', 2016, pp. 1025–1034.

[3] S. Aslam, 'Twitter by the Numbers (2020): Stats and Demographics', 05-Jan-2020. [Online]. Available: https://www.omnicoreagency.com/twitter-statistics/. [Accessed: 08- Mar-2020].

[4] H. Bagheri and M. J. Islam, 'Sentiment analysis of twitter data', ArXiv171110377 Cs, Dec. 2017. [5] H. Kvamme, 'Gender prediction on Norwegian Twitter accounts', 119, 2015.

[6] V. A. K and S., 'Sentiment Analysis of Twitter Data:A Survey of Techniques', Int. J. Comput. Appl., vol. 139, no. 11, pp. 5–15, 2016.

[7] A. &. I. Farzindar, Natural Language Processing for Social Media, 2nd Revised. San Rafael, United States: Morgan & Claypool Publishers, 2017.

[8] S. Aleksandr and T. L. D. G. R. R. I. M, 'Machine Learning Models of Text Categorization by Author Gender Using Topic-Independent Features', in 5th International Young Scientist Conference on Computational Science, 2016, vol. 101, pp. 135–142.

[9] S. Harispe and R. S. J. S. a, Semantic Similarity from Natural Language and Ontology Analysis. s.l.:Morgan & Claypool, 2015.

[10] Kaggle, Twitter User Gender Classification. 2016.

[11] F. Rangel, P. Rosso, M. Potthast, and B. Stein, 'Overview of the 5th Author Profiling Task at PAN 2017: Gender and Language Variety Identification in Twitter', p. 26.

[12] A. Bacciu, M. La Morgia, E. Nemmi, V. Neri, A. Mei, and J. Stefa, Bot and Gender Detection of Twitter Accounts Using Distortion and LSA Notebook for PAN at CLEF 2019. 2019.

[13] C. N. and C. R. S. K. P, 'Author Gender Identification from Text', Digit. Investig. Digit. Investig., vol. 8, no. 1, pp. 78–88, 2011.

[14] 'American Psychologist', Am. Psychiatr. Assoc., vol. 67, no. 1, pp. 10–42, 2012.

[15] M. Koppel and S. A. A. R. S, 'Automatically categorizing written texts by author gender', Lit. Linguist. Comput., vol. 17, no. 4, pp. 401–412, 2002.