SECJ3563 - COMPUTATIONAL INTELLIGENCE

MINI PROJECT PROPOSAL

TITLE
Fake News Detection Using Support Vector Machine (SVM)

Group Number: 4

Group Members Name	**Matric No.**
Muhammad Zhahir bin Azhar	A22EC0222
Norikwan Haikal bin Mohammed Raszainey	A22EC0236
Danish Hakimi bin Azril Dahari	A22EC0153
Luqmanul Hakim bin Mazlan	A22EC0189
Muhammad Alif Izuddin bin Azman	A22EC0202

Table of Contents

Introduction

As more aspects of our lives move online, social media platforms have become a primary source of news for many people, gradually replacing traditional news outlets. This shift in news consumption is largely driven by the characteristics of social media where the first one being it offers a more immediate and cost-effective way to access news compared to traditional mediums like newspapers and television and it facilitates easy sharing, commenting, and discussion with friends or other users (Shu et al., 2017). For instance, in 2016, 62% of U.S. adults reported getting their news from social media, which is an increase from 49% in 2012. Additionally, social media has now surpassed television as the leading source of news.

Despite the advantages offered by social media, the overall quality of news disseminated through these platforms is often inferior to that produced by traditional news organizations. This decline in quality is largely attributed to the low cost and high speed at which information can be published and shared online (Shu et al., 2017). As a result, social media has become a fertile ground for the proliferation of fake news which translates to concerns regarding the role of newswires in accelerating the unchecked dissemination of misinformation (Lazer et al., 2018).

The mini project's utilization of the Fake and Real News dataset enables the detection and reduction of the spread of misinformation across digital platforms. The dataset comprises over 44,000 news articles, each containing information such as the article's title, content, and its classification as either real or fake. The Support Vector Machine (SVM) algorithm is employed to analyze the dataset, uncover patterns within textual content, and differentiate between genuine and deceptive news articles.

Objective

The objectives of this mini project called **Fake News Detection Using Support Vector Machine (SVM)** are as follows:

- To design a machine learning model that can classify news articles as **fake** or **real**.
- To preprocess and vectorize text data effectively using **TF-IDF** techniques.
- To implement and evaluate the **Support Vector Machine (SVM)** algorithm for the classification task.
- To analyze the performance of the model using appropriate metrics such as accuracy, precision, recall, and F1 score.

Problem Background

In recent years, the growing availability of high-performance computing resources such as faster processors and larger memory capacities has significantly accelerated the adoption of machine learning (ML) techniques across various fields (Portugal et al., 2017). This surge in computational power has enabled researchers to develop and experiment with a wide range of algorithms capable of learning from data, making predictions, and uncovering hidden patterns through mathematical and statistical methods. The field continues to evolve rapidly, with numerous scientific studies proposing novel models or hybrid approaches for solving real-world problems.

One particularly important application is the use of machine learning algorithms to analyze textual data for the detection of fake news. According to McGarrigle (2022), fake news refers to false information as news, stories or hoaxes created to deliberately misinform or deceive readers. From a research done by Guo et al. (2022), fact-checking and verification are time-consuming and often fall behind the pace at which false information proliferates. Thus, automated approaches using computational intelligence techniques are critical in combating the spread of misinformation.

Recent research has demonstrated the effectiveness of various machine learning models, including Support Vector Machines (SVM) that is a support vector machine. It is mainly used to solve the problem of data classification in the field of pattern recognition (Wang et al., 2023). It belongs to a supervised learning algorithm. These models can be trained on large datasets of labeled news articles to identify deceptive language cues and predict the credibility of newly encountered articles. For instance, several projects have utilized machine learning models such as SVM and other text classification algorithms to detect fake news, demonstrating strong performance in terms of accuracy, scalability, and practical implementation in real-world information filtering systems (Baarir & Djeffal, 2021).

In this project, we aim to develop a fake news detection model using the Support Vector Machine (SVM) algorithm. By training the model on a comprehensive dataset of real and fake news articles, the system can learn to distinguish between reliable and deceptive content, potentially aiding social platforms, media outlets, and end users in filtering out misinformation effectively.

Methodology

A. Text Processing

To enhance model performance, the dataset's textual data is cleaned and preprocessed. This requires a few crucial actions. To cut down on unnecessary information, noise like special characters, URLs, and digits is first eliminated. After that, every text is changed to lowercase to maintain consistency throughout the dataset. Tokenization, which divides the text into discrete words or tokens, improves the model's ability to identify word patterns. To concentrate on more informative phrases, stopwords—common words that don't have any meaning—like "the" and "is"—are eliminated. Last but not least, normalization ensures semantic consistency throughout the data by substituting standard forms for slang phrases and repeated letters (Deokate, 2019).

B. Feature Extraction

To turn text input into numerical characteristics, we use the TF-IDF (Term Frequency-Inverse Document Frequency) approach. The SVM's capacity to identify contextual cues pertinent to the categorization of misinformation is improved by TF-IDF, which assists in quantifying the significance of words in each document in relation to the entire corpus (Gilda, 2017 as quoted in Deokate, 2019).

C. Support Vector Machine (SVM) Classification

An SVM classifier is fed the processed data. The goal of SVM is to identify the appropriate hyperplane for separating the feature vectors of authentic and fake news items. The effectiveness of this approach in binary classification problems, especially in high-dimensional spaces like text data, led to its selection (Deokate, 2019).

D. The following metrics are used to assess the SVM model when the dataset is split into training and testing sets:

E. Model Evaluation

- Accuracy: the model's overall correctness.
- Precision : the percentage of all projected fake news that turns out to be accurate.
- Recall : the percentage of real fake news that was accurately recognized.
- F1-Score : the precision and recall harmonic mean.

The assessment makes sure the model does a good job of reducing false positives and negatives in addition to accurately identifying fake news.

Conclusion

In conclusion, this project's target is to enhance awareness on the growing issue of misinformation on digital platforms. This mini project aims to tackle the misinformation on digital platforms by developing a fake news detection model using the text processing, feature extraction and Support Vector Machine (SVM) algorithm by using systematic approach involving data preprocessing. By using this approach, the project seeks to differentiate between real and fake news articles. The proposed methodology leverages computational intelligence to improve accuracy and create a reliable system in news verification. All in all, this project highlights the importance of automated tools in fighting fake news and contributes to broader efforts to ensure the credibility of online information.

Beside that, this study provides an actual illustration of how traditional machine learning methods may still deliver competitive performance in current text classification issues by concentrating on the use of TF-IDF for feature extraction and SVM for classification. The effectiveness of this method can be used as a foundational model for more complex deep learning solutions, providing information on how various algorithms behave and understand huge amounts of text data.

Furthermore, this mini project serves as a baseline for future enhancements, where state of the art approaches such as deep learning models can be integrated to boost performance and capture more sophisticated semantic meanings of the text. The current approach demonstrates the feasibility and performance of using traditional machine learning algorithms for real world fake news detection, especially in environments with limited computational resources. By establishing a solid baseline with SVM and TF-IDF, the project opens up avenues for comparative research and hybrid model development, ultimately culminating in a more robust and scalable solution to counter the increasing threat of misinformation in the age of the internet.

References

Baarir, N. F., & Djeffal, A. (2021). Fake News Detection Using Machine Learning. *2020 2nd International Workshop on Human-Centric Smart Environments for Health and Well-being (IHSH)*, 125–130. https://doi.org/10.1109/ihsh51661.2021.9378748

Deokate, S. B. (2019). *Fake News Detection using Support Vector Machine Learning Algorithm*. International Journal for Research in Applied Science & Engineering Technology (IJRASET), 7(VII), 438-444. http://doi.org/10.22214/ijraset.2019.7067

Guo, Z., Schlichtkrull, M., & Vlachos, A. (2022). A survey on Automated Fact-Checking. *Transactions of the Association for Computational Linguistics*, *10*, 178–206. https://doi.org/10.1162/tacl_a_00454

Lazer, D. M. J., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., Metzger, M. J., Nyhan, B., Pennycook, G., Rothschild, D., Schudson, M., Sloman, S. A., Sunstein, C. R., Thorson, E. A., Watts, D. J., & Zittrain, J. L. (2018). The science of fake news. *Science*, *359*(6380), 1094–1096. https://doi.org/10.1126/science.aao2998

McGarrigle, J. (2022, July 22). *Explained: What is Fake news? | Social Media and Filter Bubbles*. Webwise.ie. https://www.webwise.ie/teachers/what-is-fake-news/

Portugal, I., Alencar, P., & Cowan, D. (2017). The use of machine learning algorithms in recommender systems: A systematic review. *Expert Systems With Applications*, *97*, 205–227. https://doi.org/10.1016/j.eswa.2017.12.020

Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake news detection on social media. *ACM SIGKDD Explorations Newsletter*, *19*(1), 22–36. https://doi.org/10.1145/3137597.3137600

Wang, Y., Lin, H., Li, C., She, L., Sun, L., & Wang, J. (2023). Network Autonomous Learning Monitoring System Based on SVM Algorithm. *icWCSN 2023*, 98–102. https://doi.org/10.1145/3585967.3585984