Use of GenAI tools

The University of Exeter is committed to the ethical and responsible use of Generative AI (GenAI) tools in teaching and learning, in line with our academic integrity policies where the direct copying of AI-generated content is included under plagiarism, misrepresentation and contract cheating under definitions and offences in [TQA Manual Chapter 12.3.](#) To support students in their use of GenAI tools as part of their assessments, we have developed a category tool that enables staff to identify where use of Gen AI is integrated, supported or prohibited in each assessment. This assessment falls under the category of ***AI-supported.*** *This is because if they are used wisely, they can improve your understanding of the subject and the quality of your work.*

[You can find further guidance on using GenAI critically, and how to use GenAI to enhance your learning, on Study Zone digital.](#)

When submitting your assessment, you must include the following declaration, ticking all that apply:

AI-supported/AI-integrated use is permitted in this assessment. I acknowledge the following uses of GenAI tools in this assessment:

- ❑ *I have used GenAI tools for developing ideas.*
- ❑ *I have used GenAI tools to assist with research or gathering information.*
- ☑ *I have used GenAI tools to help me understand key theories and concepts.*
- ❑ *I have used GenAI tools to identify trends and themes as part of my data analysis.*
- ☑ *I have used GenAI tools to suggest a plan or structure for my assessment.*
- ☑ *I have used GenAI tools to give me feedback on a draft.*
- ❑ *I have used GenAI tool to generate image, figures or diagrams.*
- ❑ *I have used GenAI tools to proofread and correct grammar or spelling errors.*
- ❑ *I have used GenAI tools to generate citations or references.*
- ❑ *Other: [please specify]*
- ❑ *I have not used any GenAI tools in preparing this assessment.*

I declare that I have referenced use of GenAI outputs within my assessment in line with the University referencing guidelines.

Please note: Submitting your work without an accompanying declaration, or one with no ticked boxes, will be considered a declaration that you have not used generative AI in preparing your work

If a declaration sheet cannot be uploaded as part of an assignment (i.e. at the start of an essay), students understand that by submitting their assessment that are confirming they have followed the assessment brief and guidelines about GenAI use.

Leveraging Support Vector Machines for Accurate Fake News Detection

Diya Paul

1 Introduction

In an information age, where social media is harnessed to spread news and different perspectives on a variety of macro-level topics including politics, war, and also micro-level topics such as makeup, fashion, and entertainment, it is essential to ensure that facts are verified before dissemination. Misinformation, with social media as a catalyst, propagates at lightning speed and poses significant risks such as boycotts, cancel culture, and societal polarization. Furthermore, it can spark riots, incite mass hate, and undermine public trust in institutions.

PWC [2] categorised fake news into 3 categories; misinformation, disinformation, and malinformation. Misinformation being the sharing of incorrect info, disinformation being the intentional fabrication of falsified info and distribution of it in order to deceive of manipulate views, and malinformation which constitutes genuine information twisted or used out-of-context in order to maliciously to harm a person, organization, or country. The purpose of creating this SVM model is to be able to automatically discern whether news is false or not.

2 Methodology

The dataset chosen was from Kaggle; it's a CSV file updated weekly with a usability rating of 10.0 [1] and substantial in data volume with 72134 unique values. The dataset being large meant that there was sufficient info to train a well-versed model not prone to overfitting. The dataset had 558 rows of incomplete data; since the data is text based and not numerical evidently no mean can be imputed for these hence entries without the text were dropped as the title would not be sufficient to determine whether it was fake or real news, and entries with a missing title were simply imputed with 'no title', as this is not the main focus of the models analysis.

This left us with 72095 entries left; this was a fairly balanced dataset with (35028) fake entries and (37067) true entries, but to ensure therefore no over or under x was required; however to ensure maximum accuracy when training the model we passed in the parameter weight=balanced.

Thereafter the text was cleaned; this involved case folding to lowercase and removing any non-ASCII characters (punctuation, numbers, special characters). Then the NLTK stop word set was used to remove any unnecessary words that would not add value and unnecessary overhead. However, as I believe negation words are an important part of NLP I removed these words from the stop-word set. The last step involved lemmatization; this was chosen over stemming to ensure accuracy and meaning being kept in words. Then TF-IDF was used for feature weighting, to determine the relative importance of each individual term in the document via the frequency of occurrence and distribution of words.

The data was then trained under a linear and non-linear RBF based kernel for comparison using a standard training-validation split of 70:30. Following this a comparative analysis was completed based on 4 performance metrics: accuracy, precision, recall, and F1-Score.

3 Results

The linear kernel achieved an accuracy of 94.91% and also a precision, recall, and F1-score of 0.95 0.95 and 0.95 respectively. Therefore showcasing its ability to classify the majority of the dataset correctly, this is further supported by the computed confusion matrix. The confusion matrix revealed 19,529 correctly classified entries and 1,100 misclassified instances; with 565 false negatives (true entries misclassified as fake) and 535 false positives (fake entries misclassified as true). The non-linear kernel using RBF delivered a higher accuracy of 95.95% and also a precision, recall, and F1-score of 0.96 0.96 and 0.96 respectively. The confusion matrix revealed 20,754 correctly classified entries and 875 misclassified instances, including 488 false negatives and 387 false positives.

4 Discussion

Across the board it seems that the non-linear kernel SVM model outperforms the linear kernel SVM model. It achieved a higher accuracy of 95.95%, compared to the linear kernel's 94.91%, reflecting its enhanced ability to capture complex relationships in the data. This is reinforced by the confusion matrices with the linear kernel having a slightly higher misclassification. However, the linear kernel required 24 minutes and 34.4 seconds to train and 4 minutes and 11.9 seconds to test, while the non-linear kernel required 216 minutes and 39.6 seconds for training and 7 minutes and 0.1 seconds for testing. This substantial increase in runtime (over 7.6x the linear kernel's development time) underscores the computational overhead of the non-linear kernel, even though it delivers higher accuracy and better performance metrics.

5 Limitations

In terms of the dataset there weren't many limitations in regards to data balance, completeness. It could be argued that the dropping of 558 rows with incomplete data or imputing missing titles with 'no title' may have affected the datasets representativeness and the model's ability to generalise; but given how vast the dataset is I believe this to be insignificant.

One limitation to note is the lack of metadata on the news articles; it is unknown what their origin is (publisher, author), the date of publication, source credibility etc. These features could potentially improve the model and provide more insight to other key factors when determining if news is fake or not. It could also be used to potentially classify what kind of fake news it is.

Another potential issue could be the data source bias; even though this comes from Kaggle it may not be representative all industries, topics, or geographical locations, which could limit the model's applicability. There is also the issue of news being subjective; as the data was manually labelled it may have bias. To highlight another limitation is the binary classification which may be oversimplifying the nature of news content; partially true, opinion-based, entirely true, false etc. which could provide a more nuanced insight.

There were computational limitations which impacted my methodology as the initial approach would've involved using a CNN or LTSM for the NLP SVM model constructed. In the chosen SVM model limitations also exists in terms of its scalability and dependency of the preprocessing and feature selection. A criticism of my preprocessing involves not checking for redundant entries; duplicate or extremely similar articles, this may artificially inflate the performance of my model. Furthermore this model is limited to only the English language; which whilst it is one of the most universal languages, limits the applicability of the model to all news that exists.

6 Conclusion

The implemented SVM models demonstrate robust performance in the classification of fake and true news. However, given the substantial difference in computational power and processing time required the linear kernel may be more suited for future research. It is also recommended for future efforts to be focused on experimenting with advanced NLP methods such as BERT, CNN, or LSTMs, where computational requirements are in place, in order to review more scalable models. Alternatively efforts can also be made to explore ways to optimise the computational requirements for the scalability of this SVM model.

References

[1] Fake News Classification — kaggle.com. https://www.kaggle.com/datasets/saurabhshahane/fake-news-classification/data. [Accessed 03-12-2024].

[2] John Riccio and Amy Gibbs. The fake news problem (and what to do about it), Aug 2018.

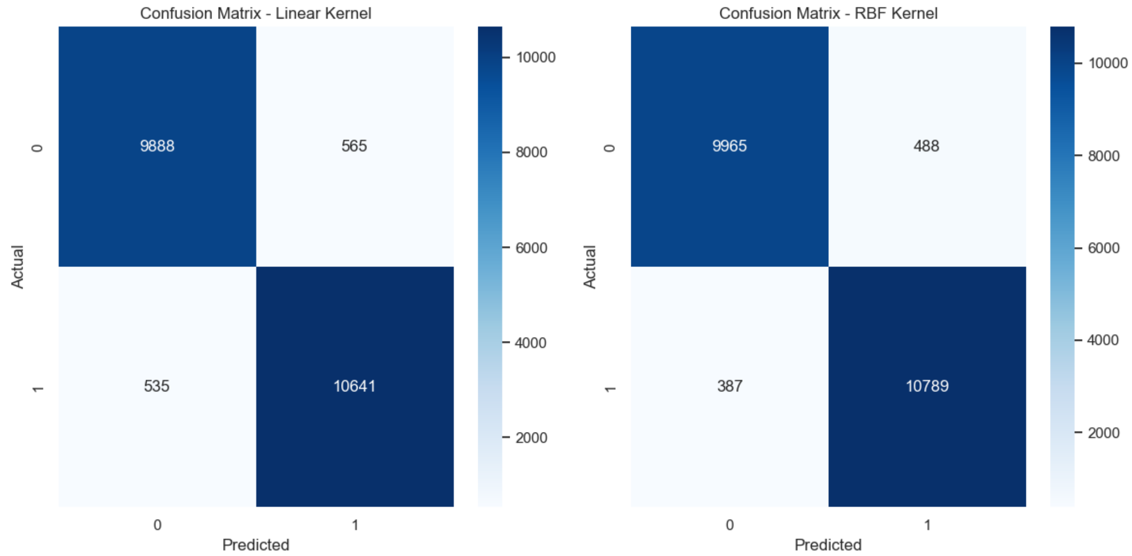

Figure 1: The figure shows the comparison of the 2 confusion matrices of the SVM models.

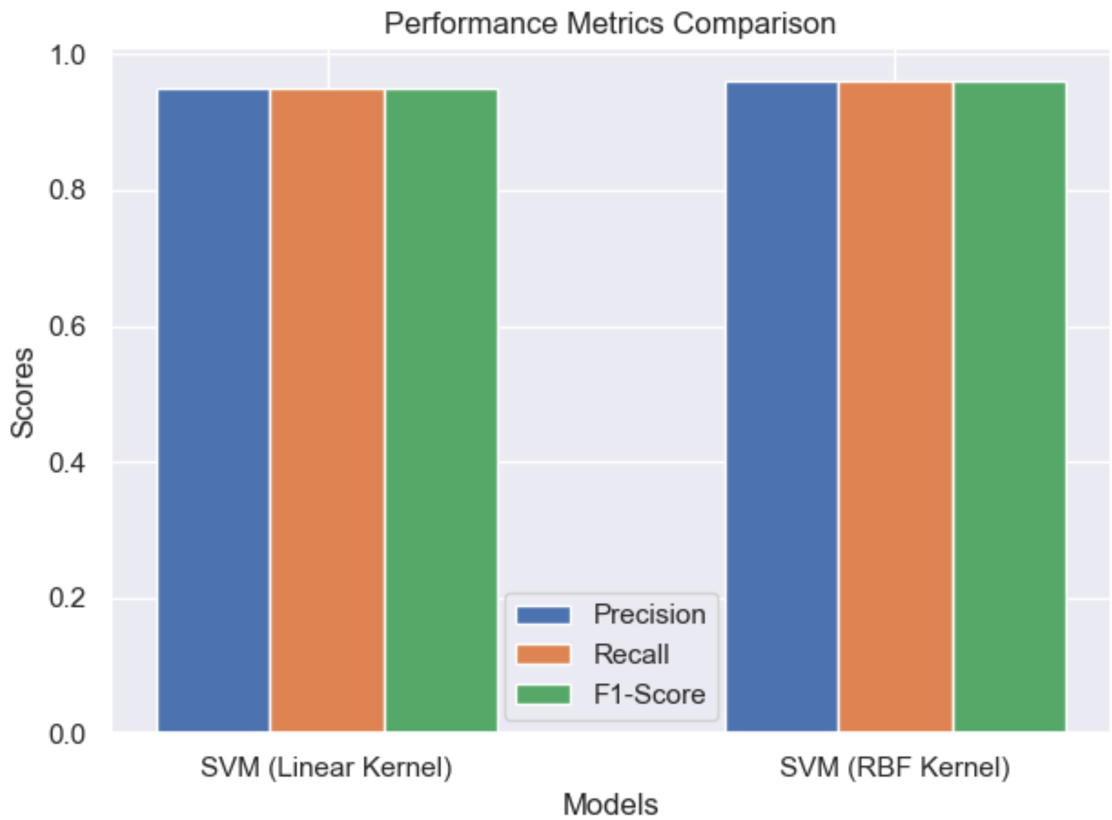

Figure 2: This figure shows the negligible difference between the models trained on the linear & non-linear kernel. Taking the processing time into consideration it is clear that the trade off between a negligible increase in accuracy of the non-linear model is not worth the computational time required.