Leveraging Support Vector Machines for Accurate Fake News Detection

A Data-Driven Approach to Combating Misinformation

By Diya Paul

Contents:

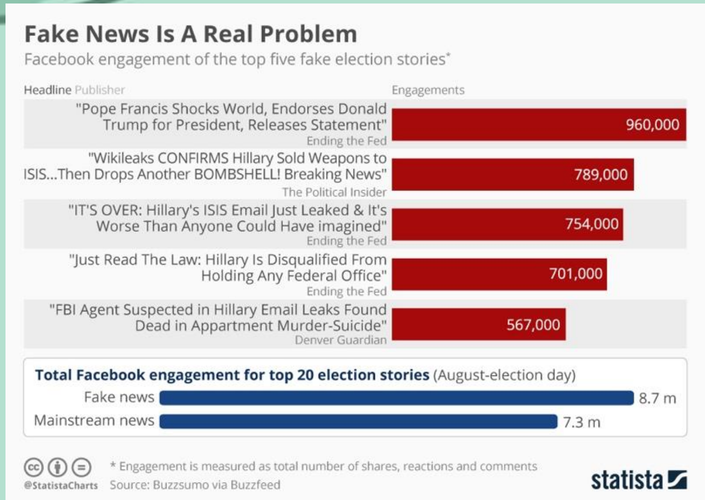
- The issue of Fake News
- Overview on SVM models
- Dataset Introduction
- Implementation Pipeline
  - Pre-processing
  - Training phase
  - Testing phase
- Results

Fake News

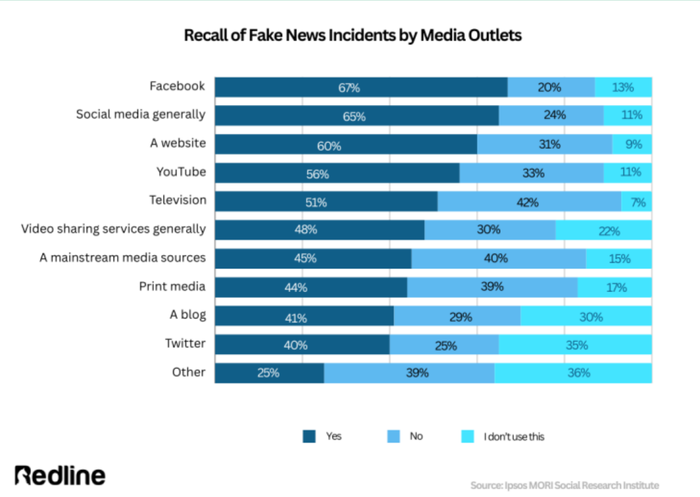
- 60% globally say news organizations regularly report false stories. [1]
- 3x increase in video deep fakes and 8x increase in voice deep fakes from 2022 to 2023. [1]

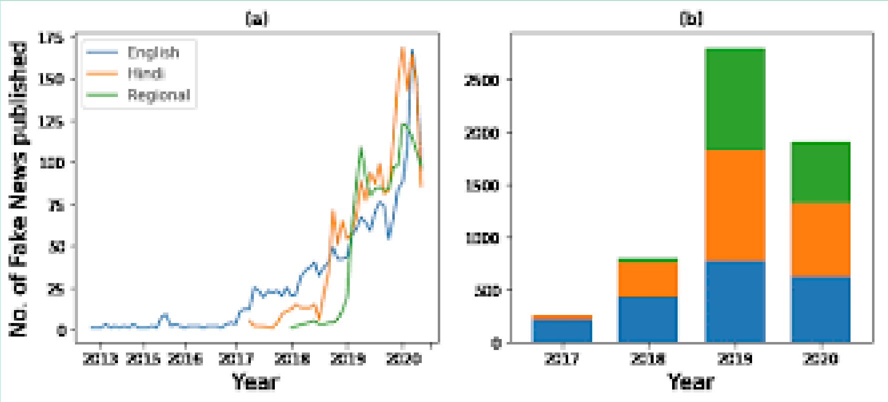
[2]

Fake News



[5] [6]





[7]

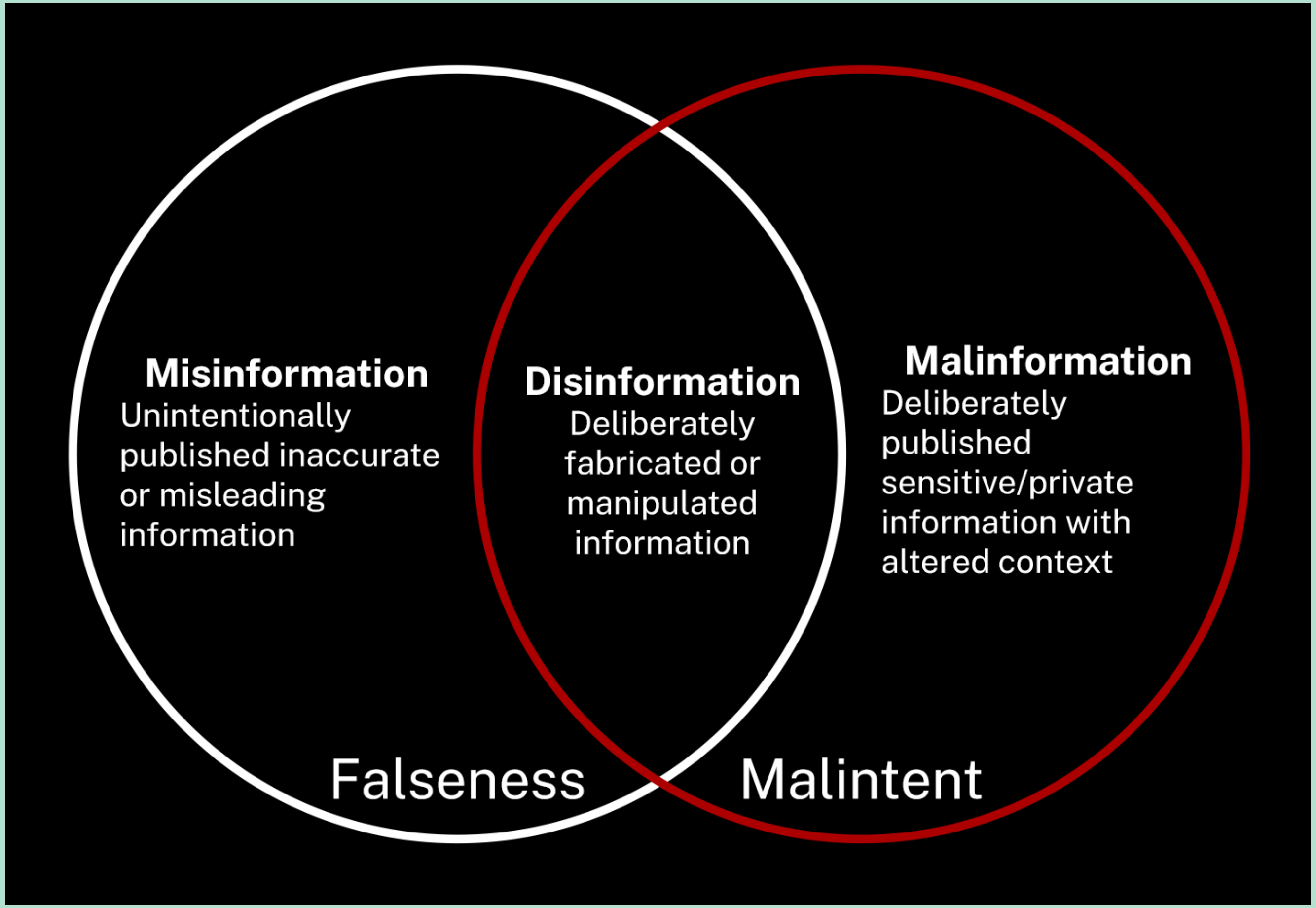
[2]

Fake News

1. **Erosion of Trust:**
    1. Undermines trust in media, institutions, and public figures.
    2. Creates scepticism about legitimate news.
2. **Social Polarization:**
    1. Deepens divides by spreading biased or false narratives.
    2. Fuels tensions between political, social, and cultural groups.
3. **Real-World Consequences:**
    1. Can incite violence, riots, and public unrest.
    2. Disrupts democratic processes (e.g., elections) through misinformation.

[2]

The classification of Fake News:



[3]

The Chosen Dataset

Source: Kaggle's Fake & Real News Dataset.

Entries: 72,134

Entries after preprocessing: 72,095

Class Distribution: Balanced (~50% fake, 50% real).

Count of True and Fake News:

1   37067

0   35028

Therefore the class distribution of Real : Fake was: 51:49.

	title	text	label
0	LAW ENFORCEMENT ON HIGH ALERT Following Threat...	No comment is expected from Barack Obama Membe...	1
1	NaN	Did they post their votes for Hillary already?	1
2	UNBELIEVABLE! OBAMA'S ATTORNEY GENERAL SAYS MO...	Now, most of the demonstrators gathered last ...	1
3	Bobby Jindal, raised Hindu, uses story of Chri...	A dozen politically active pastors came here f...	0
4	SATAN 2: Russia unvelis an image of its terrif...	The RS-28 Sarmat missile, dubbed Satan 2, will...	1

Preprocessing Pipeline

Null Values

- Dropped if in the 'text' column
- If the title column it was replaced with 'no title'

Case folding

Removal non ASCII

Stop word removal

- negation words were omitted for the stop word removal set

Lemmatisation

TF-IDF

Overview of SVM Models

SVM model is a supervised learning model typically used for classification or regression.

It does this by finding the best hyperplane that separates data points from different classes.

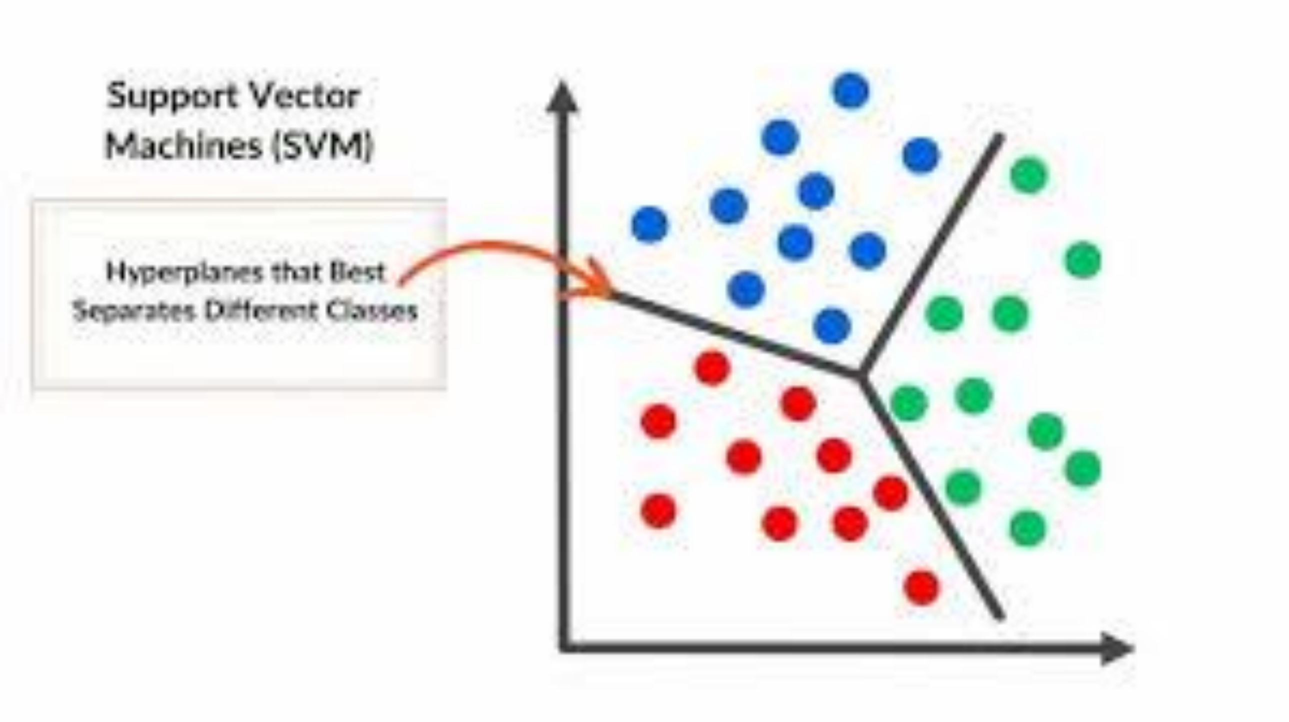
Support Vectors: the data points that are closest to the hyperplane and influence its position and orientation.

Suitable for classification as they are effective with dealing with high dimensional data & are robust to overfitting.

This was chosen as its simple to implement but powerful. We experiment with both linear & non-linear SVMs

Linear = Classifies linearly separable data.

Non-linear = Uses kernel's to map data into a higher dimensional space for non-linear separations



[4]

Training Phase

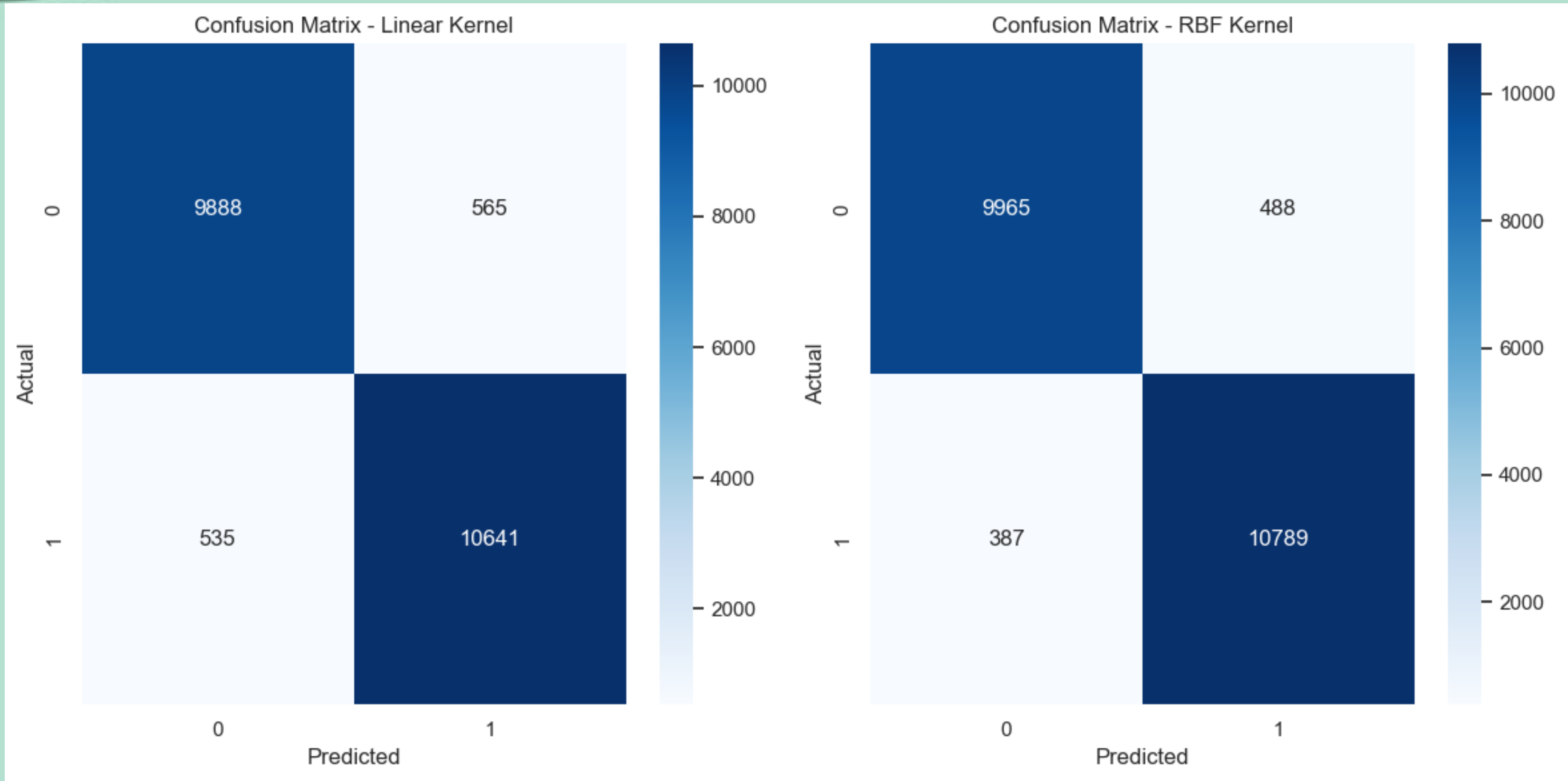
A 70:30 train-test split was used on the pre-processed data

- Total training time:
  - Linear Kernel $\Longrightarrow$ 24m 34.4s
  - Non-linear Kernel $\Longrightarrow$ 216m 39.1s
- The parameter weight='balanced' was passed as a parameter:
  - To avoid any sort of class imbalance which could cause the under or overfitting of the model.
- Random seed selected to ensure reproducibility of results and consistency in train-test split.

Testing Phase

- Total testing time:
  - Linear: 4m 11.9s
  - Non-Linear: 7m 0.1s
- Evaluation metrics:
  - Accuracy
  - Precision
  - Recall
  - F1-Score
- A Confusion Matrix was computed for each model

Results & Insights

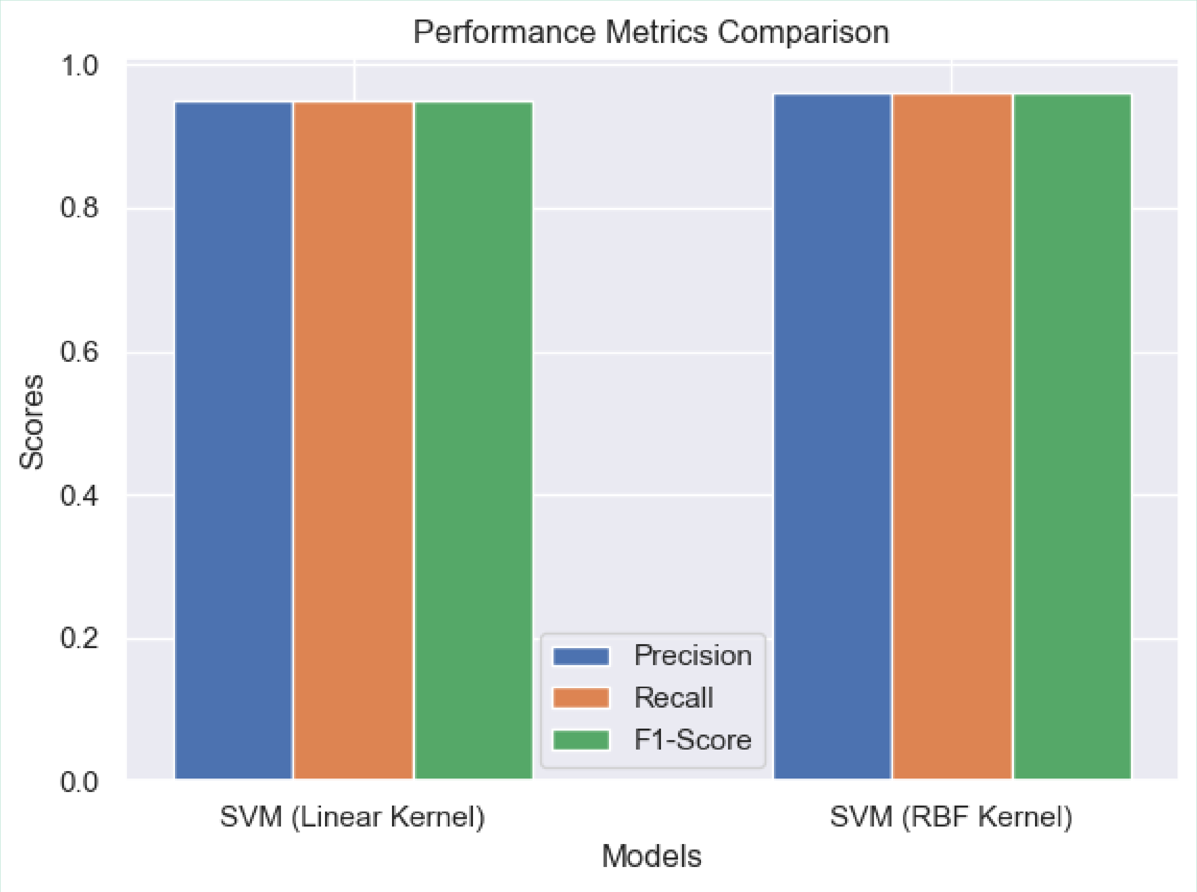


The Linear kernel had higher misclassification than the non-linear kernel

Results & Insights

The precision, recall & F1 score of the Linear kernel = 0.95 for all

Accuracy: 94.91%



The precision, recall & F1 score of the non-linear RBF kernel =0.96

Accuracy: 95.95%

Results & Insights

- Key findings:
  - Non-Linear model outperforms the model in regards to all the performance metrics
  - The run-time difference is the key differentiator:
    - Non-linear kernel had a substantial runtime over 7.6x the runtime of the linear kernel.
    - This runtime means an increased computational overhead
    - This outweighs the marginal increase in performance metrics
  - The linear kernel is recommended.

Dataset Limitations

- Minimal issues with data balance and completeness.
- Dropping 558 rows of incomplete data or imputing missing titles with 'no title' might affect representativeness
  - but deemed insignificant due to the dataset's size.
- Lack of metadata (e.g., publisher, author, publication date, source credibility)
  - limits deeper insights and potential classification of fake news types.

Data Source Bias

- Kaggle dataset may not represent all industries, topics, or geographical regions.
- Subjectivity in news content introduces potential bias
  - As manually labelled.
- Binary classification oversimplifies the complexity of news
  - News can be partially true, opinion-based, entirely true, or false etc.

Computational Limitations

- Computational limitations prevented using advanced NLP models (e.g., CNN, LSTM).
- SVM model scalability
- SVM models dependency on preprocessing and feature selection are notable constraints.
- Preprocessing criticism: Redundant (duplicate) entries were not addressed, potentially inflating model performance.
- Model only support English
  - Whilst one of the most worldwide languages the applicability is limited to non-English news.

Key Takeaways / Conclusion:

- Whilst the performance metrics of the non-linear model all around surpasses the linear model, the trade-off between computational power and time and accuracy is negligible
- Therefore linear model is the best overall.

Key Takeaways / Conclusion:

- Recommendations for future work:
  - Efforts be placed in more complex models / NLP techniques such as:
    - CNN
    - LTSM
    - BERT
  - Incorporating metadata features which may allow the classification of what type of false news an article may be
  - Utilisation of a more granular categorisation method instead of a binary method.

References

[1] Konopliov, A. (2024). *Fake News Statistics & Facts (2023) — Redline Digital*. [online] redline.digital. Available at: https://redline.digital/fake-news-statistics/.

[2] Google.com. (2024). *Redirect Notice*. [online] Available at: https://www.google.com/url?sa=i&url=https%3A%2F%2Fwww.istockphoto.com%2Fphotos%2Ffake-news-stamp&psig=AOvVaw1I_sWXQtXkMk8TkVlC5h9t&ust=1733488140682000&source=images&cd=vfe&opi=89978449&ved=0CBQQjRxqFwoTCPjEgavSkIoDFQAAAAAdAAAAABAE [Accessed 6 Dec. 2024].

[3] Google.com. (2024). *Redirect Notice*. [online] Available at: https://www.google.com/url?sa=i&url=https%3A%2F%2Flibrary.csi.cuny.edu%2Fmisinformation&psig=AOvVaw3yHAJTQDFWsg3wQVofH9nx&ust=1733488559251000&source=images&cd=vfe&opi=89978449&ved=0CBQQjRxqFwoTCPihhKDSkIoDFQAAAAAdAAAAABAE [Accessed 6 Dec. 2024].

[4] Google.com. (2024). *Redirect Notice*. [online] Available at: https://www.google.com/url?sa=i&url=https%3A%2F%2Fspotintelligence.com%2F2024%2F05%2F06%2Fsupport-vector-machines-svm%2F&psig=AOvVaw3eP2BSv5Ub7cB16w7pF8HR&ust=1733488436784000&source=images&cd=vfe&opi=89978449&ved=0CBcQjhxqFwoTCJjgho7SkIoDFQAAAAAdAAAAABAE [Accessed 6 Dec. 2024].

References

[5] Google.com. (2024). *Redirect Notice*. [online] Available at: https://www.google.com/url?sa=i&url=https%3A%2F%2Fwww.statista.com%2Fchart%2F6795%2Ffake-news-is-a-real-problem%2F&psig=AOvVaw12cBBJsRKsQ-eOumR1WQzD&ust=1733547162210000&source=images&cd=vfe&opi=89978449&ved=0CBcQjhxqFwoTCOCPjsiskooDFQAAAAAdAAAAABAE [Accessed 6 Dec. 2024].

6] Google.com. (2024). *Redirect Notice*. [online] Available at: https://www.google.com/url?sa=i&url=https%3A%2F%2Fredline.digital%2Ffake-news-statistics%2F&psig=AOvVaw1nLAkrSF05uePZEEohpZKe&ust=1733547234104000&source=images&cd=vfe&opi=89978449&ved=2ahUKEwii4P7prJKKAxXiY0EAHV7UPLwQjhx6BAgAEBo [Accessed 6 Dec. 2024].

[7] Factorization of Fact-Checks for Low Resource Indian Languages - Scientific Figure on ResearchGate. Available from: https://www.researchgate.net/figure/Circulation-of-fake-news-over-the-years-in-India-For-the-year-2020-the-data-is_fig5_349547341 [accessed 6 Dec 2024]