Group 6: Fake News Detection

Motivation

The rapid proliferation of fake news has emerged as a significant societal challenge, exacerbated by the widespread use of social media and other online platforms. Fake news can influence public opinion, undermine democratic processes, and erode trust in credible sources of information.

Given the sophistication with which fake news can mimic credible sources, distinguishing between fake and authentic news has become increasingly difficult and essential. By developing robust machine learning models to accurately classify news articles, we can provide a critical tool for ensuring news credibility.

This will help citizens make informed decisions, support the integrity of democratic processes, and restore trust in genuine news sources. The potential impact of this research extends beyond individual use cases, contributing to broader efforts to combat misinformation and protect the public from deceptive content.

Concise Problem Definition

How can we effectively distinguish between fake and authentic news stories using machine learning techniques?

The problem is novel and interesting for the following reasons:

1. **Data Imbalance**: Alongside comparing the performance of different machine learning models, we are also evaluating and comparing two approaches to address data imbalance: augmenting the dataset by collecting additional data and applying sampling techniques.

2. **Variety of Features**: This problem involves analyzing a diverse range of features, including linguistic patterns, source credibility, and contextual information, making it a comprehensive and multifaceted challenge.

3. **Technological Relevance and Model Comparison**: Advances in machine learning and natural language processing provide new opportunities to tackle this problem effectively. We compare the effectiveness of basic classification models with advanced models.

4. **Societal Impact**: Successfully addressing this problem can have significant societal benefits by reducing the spread of misinformation and mitigating its negative consequences.

Related work

In 2015 November, Conroy, N. J., Rubin, V. L., & Chen, Y in their research paper described Linguistic Cue Approaches with machine learning, Bag of words approach ,Rhetorical Structure and discourse analysis ,Network analysis approaches and SVM classifiers. These models are text based only.

"Fake News Detection on Social Media: A Data Mining Perspective", Authors: Shu, Kai, et al.,. This paper provides a comprehensive overview of fake news detection techniques from a data mining perspective. It explores various approaches, including content-based analysis, social context analysis, and propagation pattern analysis, for identifying and combating fake news on social media platforms.

In 2015, November, Chen, Y., Conroy, N. J., & Rubin, V. L have described Tabloidization in the form of Click baiting. They have described Click baiting as a form of rapid dissemination of rumor and misinformation online . The authors have discussed potential methods for automatic detection of clickbait as a form of deception. Content cues which include lexical and semantic level of analysis were implemented by the authors.

In 2017, Natali Ruchansky, Sungyong Seo, and Yan Liu used a hybrid network, merging news content features and metadata such as social engagement in a single network. To do so, they used an RNN for extracting temporal features of news content and a fully connected network in the case of social features. The results of the two networks are then concatenated and used for final classification. As textual features they used doc2vec. They did test their model on two datasets, one from Twitter and the other one from Weibo, which is a Chinese equivalent of Twitter. Compared to simpler models, CSI performs better, with 6% improvement over simple GRU networks.

In 2017, Mykhailo Granik, Volodymyr Mesyura in their research paper concluded their approach for fake news detection using Naive Bayes classifier which has presented an accuracy of 74% on the test set.

Data

The study utilizes the Fake and Real News Dataset from Kaggle (https://www.kaggle.com/datasets/clmentbisaillon/fake-and-real-news-dataset/data), which was compiled by merging two existing datasets: BS Detector and Fake News Corpus. The dataset contains the following characteristics:

- 44,898 news articles in English
- Articles labeled as "FAKE" or "REAL"
- Covers various categories such as politics, sports, entertainment, etc.
- Includes news articles from a wide range of reliable and unreliable sources
- Four columns: Title, Text (body text), Subject, and Date (publish date)

This dataset is well-suited for the fake news detection problem due to the following reasons:

1. **Widely Used and Benchmarked:** The dataset is widely used in fake news detection research, allowing for comparison of results with existing literature and benchmarks. This facilitates the evaluation of model performance and contributes to the advancement of the field.
2. **Size and Diversity:** With 44,898 news articles covering a wide range of topics and sources, the dataset provides a large and diverse sample for training machine learning models. This diversity helps the models generalize better and perform well on unseen data, making them more robust and reliable in real-world scenarios.
3. **Realistic Scenario:** The dataset includes news articles from various sources, both reliable and unreliable, mimicking the real-world scenario where fake news can originate from different sources and blend with real news. This realistic representation helps in training models that can effectively distinguish fake news from real news in practical situations.

Approach

Our approach to distinguishing between fake and authentic news stories involved several key steps: addressing data imbalance, selecting, and evaluating various machine learning models, and employing rigorous evaluation criteria to ensure robust performance.

1. Addressing Data Imbalance

The original dataset was compiled from two individual datasets, each exhibiting class imbalance when used separately. To address this, we implemented two approaches:

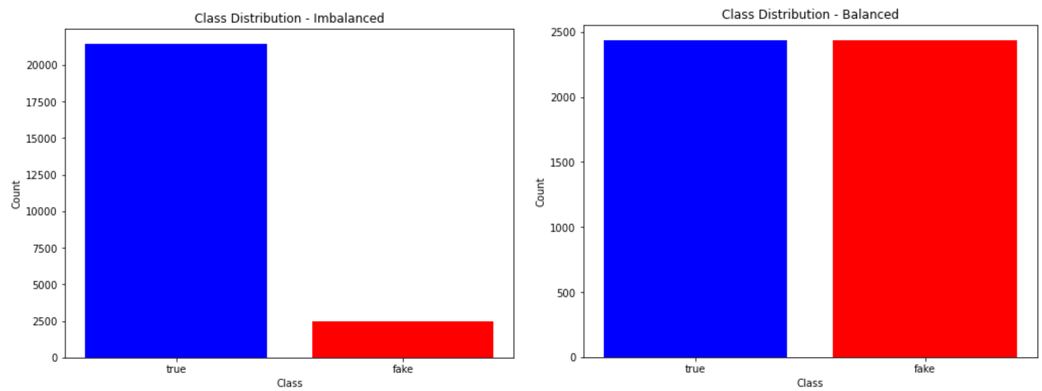
1. Using the entire balanced dataset by combining both datasets.
2. Using one of the individual imbalanced datasets, and applying sampling techniques to balance it.

We then compared the results from both approaches to evaluate the effectiveness of our methods.

Sampling Technique - Undersampling:

- **Technique**: The imbalanced dataset we used contained a higher number of true news instances compared to fake news instances. To address this, we applied the NearMiss algorithm, which reduces the number of majority class instances (true news) while retaining all instances of the minority class (fake news). This ensures that the model is not biased towards the majority class.
- **Steps**:

- **Selection Criteria**: Identifying and removing instances of the majority class based on their distance to the minority class instances.
- **Implementation**: Using NearMiss, we achieved a balanced dataset by undersampling the true news articles.



2. Model Selection and Training

We employed a variety of models to explore different approaches to the classification task. Basic models included Naive Bayes, Logistic Regression, and Support Vector Machine (SVM), chosen for their simplicity and effectiveness in text classification. Advanced models included Decision Trees, Random Forest, and Long Short-Term Memory (LSTM) networks, selected for their ability to capture complex patterns and dependencies in the data. This combination allowed us to compare the performance of simple and sophisticated techniques in distinguishing between fake and authentic news.

3. Evaluation Criteria

We evaluated our approaches and models using accuracy, precision, recall, and F1 score to assess their effectiveness in distinguishing between fake and authentic news. Among these metrics, we placed particular emphasis on the Recall metric. Ensuring that fake news is correctly identified is crucial, and Recall measures our model's ability to do so by identifying all actual fake news articles. This minimizes the misclassification of fake news as true, enhancing the overall reliability of our approach.

Validation:

- **Cross-Validation**: We employed k-fold cross-validation to ensure the robustness of our models. This involved splitting the dataset into k subsets, training the model on k-1 subsets, and testing it on the remaining subset, iteratively.
- **Confusion Matrix**: We analyzed the confusion matrix to understand the distribution of true positives, true negatives, false positives, and false negatives.

4. Implementation and Experimentation

Data Preprocessing:

- **Text Cleaning**: Removing stopwords, punctuation, and special characters to ensure the textual data is in a suitable format for analysis.
- **Feature Extraction**: Using techniques like TF-IDF (Term Frequency-Inverse Document Frequency) to convert textual data into numerical features.

Model Training:

- Each model was trained on two different datasets: the original balanced dataset and the dataset balanced using undersampling.
- **Hyperparameter Tuning**: Optimal hyperparameters were determined using grid search and random search methods.

By following this comprehensive approach, we ensured a rigorous and methodical evaluation of various machine learning techniques to distinguish between fake and authentic news stories effectively.

Results

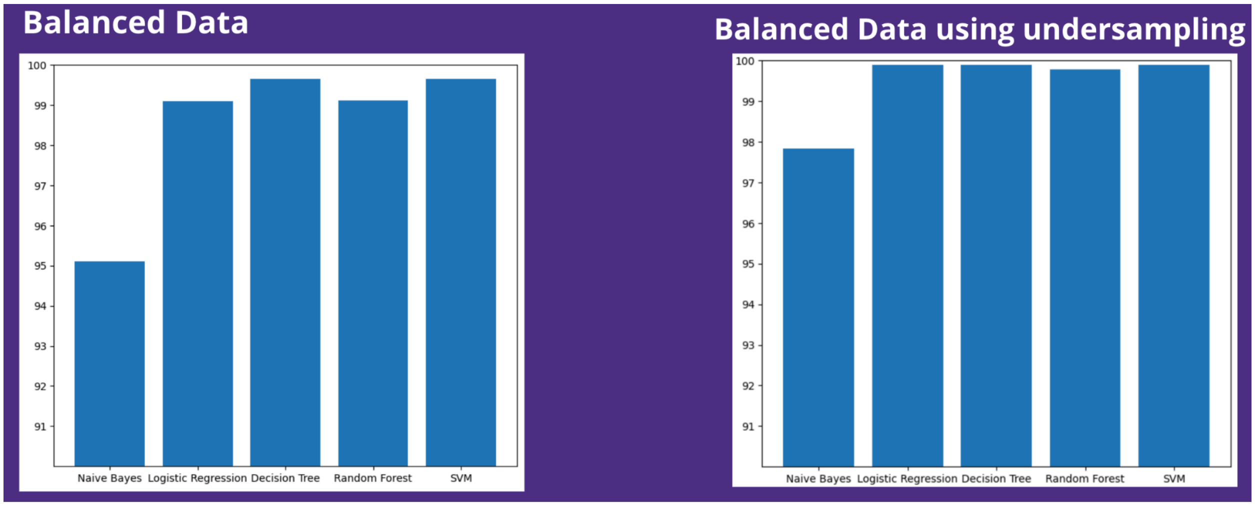
This section analyzes the impact of data balancing techniques on the performance of various machine learning models for fake news detection. Our experiment utilized two approaches:

1. Balanced Data: The initial dataset contained an equal number of data points for real and fake news articles.

2. Balanced Data using Undersampling: The majority class (likely real or fake news) was undersampled to create a balanced dataset.

Machine Learning Models:

1. Naive Bayes
2. Logistic Regression
3. Decision Tree
4. Random Forest
5. SVM



The graphs depict the accuracy achieved by each model under both balanced data and undersampled data scenarios.

LSTM Approach

Long Short-Term Memory (LSTM) neural networks are a specialized type of recurrent neural network (RNN) designed to recognize patterns and dependencies in sequential data over time. Unlike traditional RNNs, LSTMs are specifically tailored to retain information for extended periods, making them particularly effective for tasks involving long-term dependencies.

How Does LSTM Work?

The functionality of LSTMs revolves around the flow of information through a series of gates and memory cells, which collaboratively manage the addition, retention, and output of data. These components ensure that relevant information is maintained over long sequences, while irrelevant information is discarded.

LSTM Components and Information Flow

LSTM networks use gates and memory cells to manage information flow. The input gate adds significant new information to the memory cell, while the forget gate discards unnecessary data, ensuring the network remains focused on relevant details. The output gate determines which information is output or passed to the next layer. Memory cells store essential details, maintaining the network state by updating, retaining, or outputting information based on signals from the gates. The information flow involves input processing, where the input gate evaluates and adds relevant data to the memory cell; selective forgetting, where the forget gate discards unnecessary data; and output generation, where the output gate processes the memory cell state and determines the output.

LSTM Experiments

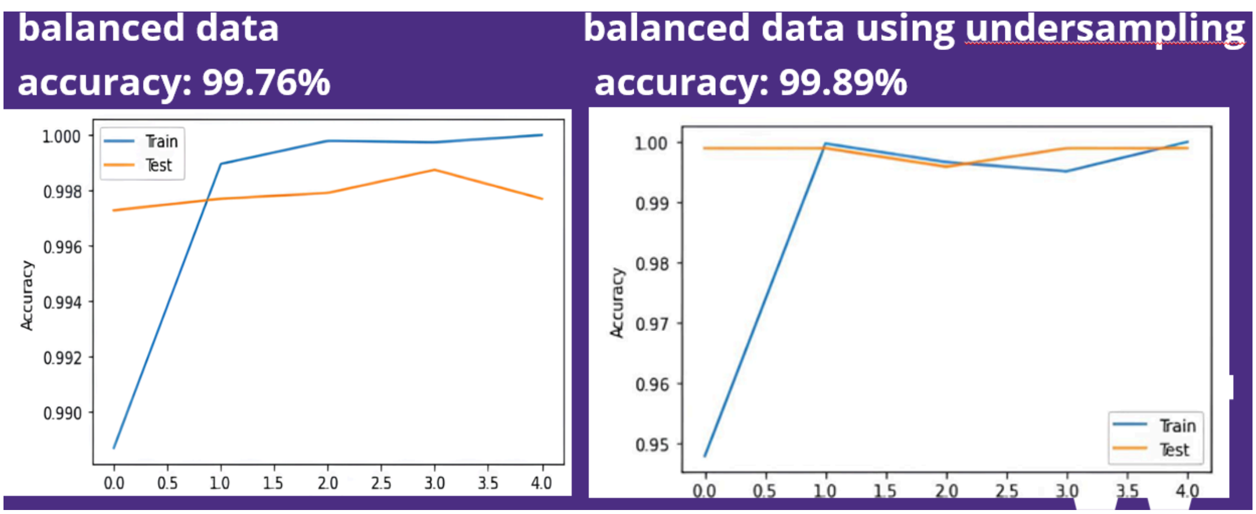
1. **Data Preparation:** The environment is set up with essential libraries: Pandas for data manipulation, NLTK for natural language processing, TensorFlow for model building, and Matplotlib for visualization. NLTK stopwords are downloaded to remove common, non-informative words from the text.

2. **Loading and Preprocessing Data:** Loading Data. Two datasets of news articles are used:'Fake' news labeled as 0.'True news labeled as 1. The datasets are concatenated and shuffled to ensure randomness. Text preprocessing includes converting text to lowercase, Removing punctuation, Filtering out stopwords using NLTK.

3. **Tokenization and Padding:** Text data is transformed into numerical sequences using TensorFlow's Tokenizer, converting words into numerical tokens. Sequences are padded using TensorFlow's `pad_sequences` to ensure uniform length, necessary for neural network training.

4. **Building the LSTM Model:** The model consists of the Embedding Layer that learns dense word representations, the LSTM Layer that Captures patterns in the text and the dense layer with Sigmoid Activation Which Enables binary classification to distinguish between 'fake' and 'true' news.

LSTM Results



LSTM results for balanced data using two methods: undersampling and an unspecified technique.

Balanced Data

- Training Accuracy: Quickly rises to near 1.000 and remains stable.
- Testing Accuracy: Gradually increases to around 0.9976, with a minor dip.

Balance Data using Undersampling

- Training Accuracy: Rapidly increases to near 1.000.
- Testing Accuracy: Starts around 1.000, dips slightly, then stabilizes around 0.9989.

Both methods achieve high training accuracy.The balanced data using undersampling technique maintains slightly higher and more consistent testing accuracy compared to the balanced data.

Findings

Impact of Undersampling on Fake News Detection Performance

Our analysis revealed interesting insights into the effect of balancing data using undersampling on the performance of various machine learning models for fake news detection.

Improved Accuracy Across Models:

Contrary to the initial expectation of similar or slightly improved performance with balanced data, undersampling led to a noticeable **increase in accuracy** for all models. This is evident in the graphs, where every model shows a positive shift in accuracy after undersampling.

The most significant improvement was observed in the Naive Bayes model, which initially had the lowest accuracy. This suggests that the model was particularly susceptible to the class imbalance present in the original dataset. Undersampling likely provided a more balanced learning environment, allowing Naive Bayes to learn the intricacies of both real and fake news with greater accuracy.

Encouragingly, models like Logistic Regression, Decision Tree, Random Forest, and SVM, which already exhibited high accuracy, also benefited from undersampling. These improvements highlight the potential of undersampling to further **optimize** the performance of even well-performing models.

Effectiveness of Undersampling:

The observed improvements can be attributed to the effectiveness of undersampling in addressing the **class imbalance** issue. By reducing the number of instances in the majority class, undersampling ensured that the models were not biased towards the dominant class. This resulted in a more **balanced learning environment**, where the models were exposed to an equal representation of both real and fake news articles.

This balanced representation likely allowed the models to learn the distinguishing features of both classes more accurately. Consequently, the models were able to **generalize** their learnings better to unseen data in the test set, leading to the observed increase in accuracy across all models.

Our findings demonstrate the **positive impact** of undersampling on the performance of machine learning models for fake news detection. Undersampling not only improved the performance of models that were initially hindered by class imbalance but also further optimized the accuracy of already well-performing models. This suggests that undersampling can be a valuable preprocessing technique for tasks involving imbalanced datasets, particularly when dealing with crucial tasks like fake news detection.

Ethical Considerations

While undersampling demonstrates promise in improving fake news detection, it's crucial to consider the ethical implications of this approach.

Bias and Fairness:

- **Diverse Dataset Representation:** Undersampling from the majority class can inadvertently remove valuable data points representing legitimate viewpoints. Ensure the remaining data accurately reflects the real world to prevent bias against specific sources or perspectives.
- **Balancing Techniques:** Carefully evaluate different balancing techniques beyond just undersampling. Techniques like SMOTE, which create synthetic data points for the minority class, might mitigate bias towards the majority class while addressing class imbalance.

Misuse Prevention:

- **Intentional Misinformation:** Undersampling might not be effective against actors who intentionally craft fake news to mimic characteristics of the minority class. Explore complementary techniques to identify and address such malicious attempts.
- **Ethical Guidelines:** Develop and adhere to ethical guidelines for fake news detection systems. These guidelines should address issues like data privacy, transparency in model decision-making, and responsible use of the technology.

Continuous Monitoring:

- **Regular Updates:** Machine learning models can become outdated as fake news tactics evolve. Regularly update the models with new data and monitor their performance to ensure continued effectiveness.
- **Adapt to New Trends:** Stay informed about emerging fake news trends and adapt the detection system accordingly. This might involve incorporating new features or retraining the model with data reflecting the latest tactics used by purveyors of misinformation.

By incorporating these ethical considerations, we can ensure that fake news detection systems using undersampling are not only effective but also responsible and fair.

Limitations and Future work

While undersampling proved beneficial in our experiment, it's important to acknowledge the inherent limitations of this approach and explore potential avenues for future work.

Limitations:

1. **Dataset Bias:** The performance improvements observed may be specific to the dataset used. Generalizability to other datasets containing different types of fake news or varying imbalances might be limited. It's crucial to evaluate the effectiveness of undersampling on diverse datasets to ensure robustness.
2. **Imbalanced Classes for Rare Instances:** Undersampling inherently reduces the number of data points in the majority class. While this improves overall accuracy, it can lead to a decrease in performance for the minority class, especially for rare instances of fake news. Techniques like SMOTE (Synthetic Minority Over-sampling Technique) that create synthetic data points for the minority class might be worth exploring to mitigate this issue.
3. **Lack of External Validation:** The current results are based on a single dataset. Validating our findings on completely external datasets would strengthen the overall conclusions about the effectiveness of undersampling for fake news detection tasks.
4. **Interpretability:** Machine learning models, especially complex ones, can be difficult to interpret. Understanding why a specific news article is classified as fake news can be

crucial. Future work should explore incorporating interpretability techniques to gain insights into the models' decision-making processes.

Future Work:

1. **Explore Different Architectures:** Investigating the impact of using different machine learning architectures, such as deep learning models, on fake news detection with balanced and imbalanced datasets could yield valuable insights.
2. **Multimodal Approach:** Fake news often incorporates multimedia elements like images and videos. Exploring a multimodal approach that incorporates text analysis alongside image and video analysis might improve detection accuracy.
3. **Transfer Learning:** Leveraging pre-trained models on large datasets for fake news detection tasks is a promising avenue for future research. This could potentially improve performance and reduce training times.
4. **Explainability Techniques:** Incorporating explainability techniques like LIME (Local Interpretable Model-Agnostic Explanations) can help understand how models arrive at their classifications, especially when dealing with complex models.
5. **Real-time Application:** Developing real-time fake news detection systems that can be integrated into social media platforms or news websites would be a significant contribution towards mitigating the spread of misinformation.

Conclusion

The learning models were trained and parameter-tuned to obtain optimal accuracy. Some models have achieved comparatively higher accuracy than others. We used multiple performance metrics to compare the results for each algorithm. The ensemble learners have shown an overall better score on all performance metrics as compared to the individual learners.

LSTM models, with and without undersampling, excel in distinguishing fake news from authentic stories, showcasing near-perfect accuracy and robustness across datasets. Both models achieve high accuracy, demonstrate robust performance, and exhibit efficient learning from data. This study contributes advanced LSTM models to fake news detection, evaluates the impact of undersampling on model performance, and establishes the reliability of LSTM models for real-world news classification.

References

[1]- https://asistdl.onlinelibrary.wiley.com/doi/full/10.1002/pra2.2015.145052010082

[2]- https://dl.acm.org/doi/10.1145/3137597.3137600

[3]- https://arxiv.org/abs/1703.06959

[4]- https://www.semanticscholar.org/paper/Fake-news-detection-using-naive-Bayes-classifier-Granik-Mesyura/248c6140213c65b631cb3847293adf3fe91d1c4d

[5]-https://www.researchgate.net/publication/283721117_Misleading_Online_Content_Recognizing_Clickbait_as_False_News