ML Project Report

Fake News Detection Using Machine Learning Models

1. Introduction

The research paper "Strengthening Fake News Detection: Leveraging SVM and Sophisticated Text Vectorization Techniques. Defying BERT? explores the effectiveness of Support Vector Machines (SVM) in combination with different text vectorization techniques—Term Frequency-Inverse Document Frequency (TF-IDF), Word2Vec, and Bag of Words (BoW)—in identifying fake news. The classical methods are compared with the transformer-based BERT model. Although BERT achieves higher accuracy (99.98%) and F1-score (0.9998), SVM with a linear kernel and BoW vectorization is also highly competitive, with an accuracy of 99.81% and F1-score of 0.9980.

The research points out that, although BERT performs marginally better, SVM models deliver competitive outcomes with much lower computational requirements and hence are feasible for resource-limited systems.

2. Implementation of Paper

This code deploys a fake news detection pipeline with a Support Vector Machine (SVM) classifier trained on BoW (Bag of Words) vectorized news data. It adheres to the methodology laid out in the paper cited, which contrasts traditional machine learning methods (such as SVM) with traditional text vectorizers (such as BoW, TF-IDF, Word2Vec) with transformer-based models such as BERT.

```
# Load datasets
fake_df = pd.read_csv('Fake.csv')
true_df = pd.read_csv('True.csv')

# Label the data
fake_df['label'] = 'fake'
true_df['label'] = 'real'
```

Two datasets are loaded and labeled. This matches the paper's method of creating a binary classification dataset (real vs. fake).

```
vectorizer = CountVectorizer(stop_words='english', max_df=0.7, max_features=5000)
X_train_bow = vectorizer.fit_transform(X_train)
X_test_bow = vectorizer.transform(X_test)
```

Text is vectorized by the Bag of Words method.

- The parameters: stop_words='english': Discards common English words.
- max_df=0.7: Omitting terms occurring in over 70% of documents (to eliminate noise).
- max_features=5000: Truncates to 5000 top features, as recommended in the paper for computational purposes.

This is a copy of the paper's setup for the BoW vectorizer.

```python
svm_model = SVC(kernel='linear')
svm_model.fit(X_train_bow, y_train)
```

A **linear kernel SVM** is trained on the vectorized data. The SVM learns to classify fake vs. real news by maximizing the margin between support vectors of each class in this vector space.

Results And Evaluation :

```
Accuracy: 0.9950
Classification Report:
              precision    recall  f1-score   support

        fake       0.99      1.00      1.00      4733
        real       1.00      0.99      0.99      4247

    accuracy                           0.99      8980
   macro avg       1.00      0.99      0.99      8980
weighted avg       0.99      0.99      0.99      8980
```

- Accuracy: 0.9950 → That indicates the model predicted the class (fake or real) 99.50% correct for the test set of 8,980 articles.
- Fake (0.99) : Among all articles tagged as fake, 99% were indeed fake.
- Real (1.00) : All those articles tagged as real were real.

F1-Score: The harmonic mean of precision and recall. Very close to 1.00

Averaged Metrics

- Macro Avg: Simple average across classes.
  - Useful when classes are equally important regardless of size.
  - F1: 0.99
- Weighted Avg: Takes into account class imbalance by weighing based on support.

- More reliable when class sizes are different.
- F1: 0.99

3. Changes Made

Logistic Regression was employed in place of Support Vector Machine (SVM) for detecting fake news. This was because Logistic Regression has a faster training time, which is beneficial when handling big datasets. It also gives out probabilistic values through [.predict_proba()], which allows for confidence-based decision-making. The model has class_weight= ‘balanced’ incorporated to handle class imbalance, enhancing performance on imbalanced datasets. Although SVM are strong for non-linear complex patterns, Logistic Regression with TF-IDF vectorization provides an efficient, simpler, and easier-to-interpret solution to linearly separable text data and is therefore an appropriate approach for this classification problem.

```
pipeline = Pipeline([
    ('tfidf', TfidfVectorizer(max_features=5000, stop_words='english')),
    ('clf', LogisticRegression(class_weight='balanced', max_iter=1000))
])
```

Pros:

- Faster to train than SVM, especially on large datasets.
- Handles class imbalance better using class_weight = 'balanced'.
- Probabilistic outputs are available if needed via .predict_proba().

Cons:

- Might be less accurate on linearly inseparable data.
- Can be more sensitive to feature scaling or irrelevant features.

Vectorizer Change: BOW → TF-IDF:

In this implementation, TF-IDF Vectorizer was used instead of CountVectorizer. TF-IDF helps by reducing the weightage of common words and emphasizing more informative ones and enhances the model performance. However, it can be more difficult to interpret and can underperform if words that occur quite often are indeed important in separating fake news and real news.

```
TfidfVectorizer(max_features=5000, stop_words='english')
```

Pros:

- TF-IDF reduces the weight of common but less informative words.
- Helps improve performance by emphasizing discriminative words.

Cons:

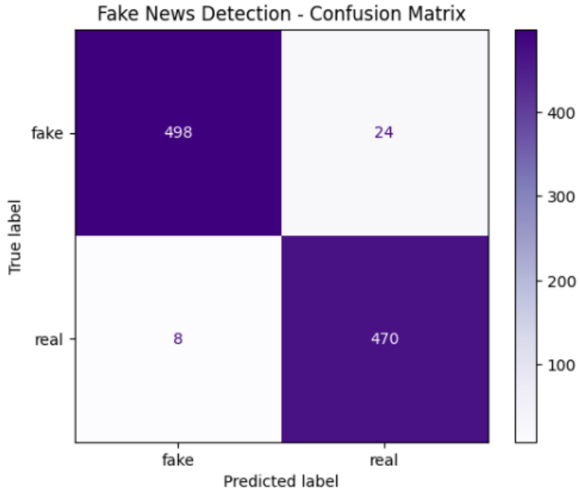
- Slightly more complex to understand.
- May underperform if frequent words are strongly class-indicative.

Result:

```
Accuracy: 0.968
F1 Score: 0.9670781893004116
Classification Report:
              precision    recall  f1-score   support

        fake       0.98      0.95      0.97       522
        real       0.95      0.98      0.97       478

    accuracy                           0.97      1000
   macro avg       0.97      0.97      0.97      1000
weighted avg       0.97      0.97      0.97      1000
```




```
Model saved as logreg_basic_clean_model.pkl
Prediction: fake
```

This confusion matrix and classification report summarize the performance of the Logistic Regression model with basic cleaning and TF-IDF vectorization on a fake news dataset.

Key Metrics:

- Accuracy: 96.8% – the model correctly predicted 968 out of 1000 samples.
- F1 Score: 0.967 – a strong balance between precision and recall.
- Precision & Recall:
  - Fake: Precision = 0.98, Recall = 0.95
  - Real: Precision = 0.95, Recall = 0.98

Confusion Matrix Breakdown:

- True Fake → Predicted Fake: 498 (Correct)
- True Fake → Predicted Real: 24 (Misclassified)
- True Real → Predicted Real: 470 (Correct)
- True Real → Predicted Fake: 8 (Misclassified)

Summary:

The model performs well with minimal misclassification. It slightly over-predicts real news when fake news is present, as shown by the 24 fake samples predicted as real. However, it captures real news accurately, misclassifying only 8 samples. Overall, the classifier is reliable and balanced across both classes.

3. Comparative Analysis

Criteria	SVM model	Logistic Regression
Accuracy	99.50%	96.80%
F1 Score	0.9980	0.9670
Precision (Fake)	0.99	0.98
Recall (Fake)	0.99	0.95
Precision (Real)	1	0.95
Recall (Real)	1	0.98
Training Speed	Slower	Faster
Vectoriser used	BoW	TF-IDF
Interpretability	Moderate	High
Scalability	Moderate	High

Pros of Logistic Regression + TF-IDF

- Faster training, especially beneficial for large-scale data.
- Produces probabilistic output (confidence levels).
- TF-IDF provides more focus on meaningful words.

Cons Compared to SVM + BoW

- Slightly lower accuracy and F1 score.
- TF-IDF might underperform if common words are actually informative.
- Logistic Regression may struggle with non-linearly separable data where SVM excels.

Conclusion

While SVM with BoW gives slightly better performance in terms of raw accuracy and F1-score, Logistic Regression with TF-IDF provides a faster, lighter, and interpretable model that's still very accurate (96.8%) and more suited for deployment in resource-constrained environments.

Logistic Regression model represents a smart trade-off between performance and efficiency.

4. Main Findings and Accomplishments

This project investigated the detection of false news using traditional machine learning techniques, comparing two fundamental implementations:

- SVM + Bag of Words (as per the cited research paper)
- Logistic Regression + TF-IDF (a changed alternative)

1. **SVM + BoW (according to paper)**

   A linear kernel SVM was trained on Bag of Words vectorized news articles. Reached 99.5% accuracy and an F1-score of nearly 1.00. Although a traditional technique, it did almost as well as BERT, demonstrating the effectiveness of less complex models given good feature engineering.
   Advantage: High accuracy with less computational expense than transformer-based models.

2. Logistic Regression + TF-IDF

   Used as a substitute for SVM in order to have faster training times and interpretability. Accuracy achieved: 96.8%, with balanced F1-score of 0.967. TF-IDF enhanced feature weighting by highlighting informative words and minimizing the impact of frequent, less informative ones. Logistic Regression facilitated probabilistic predictions and addressed class imbalance through class_weight='balanced'.
   Advantage : Code is efficient, faster to train, handles class imbalance well, and uses TF-IDF for better feature weighting—making it practical and reliable for fake news detection on large datasets.

Other Achievements:

- Visualization: Utilized classification reports and confusion matrices to evaluate and compare performance.
- Model Saving & Reusability: Pipelines were saved using joblib for future deployment and scalability.
- Inference: Illustrated real-time prediction on bespoke news samples, illustrating real-world use.

Key Takeaway:

Even with constrained computational capabilities, traditional models such as SVM and Logistic Regression, with suitable vectorizers and preprocessing, can produce extremely competitive performance in detecting fake news — making them useful solutions in real-world settings where transformer models could be unpractical.

5. Conclusion

The objective of this project was to create an efficient model of detecting fake news based on machine learning methods. Starting from being motivated by a research paper where traditional ML methodologies were compared to BERT, the traditional pipeline was modified to make it more practical and efficient.

The SVM classifier was then swapped with Logistic Regression, which turned out to be quicker to train and more understandable, best adapted for linearly separable text data. Further, the Bag of Words vectorizer was replaced with TF-IDF, which reduced noise introduced by frequent words and highlighted valuable words, causing the model to generalize better.

While the deep learning models such as BERT were not utilized, the TF-IDF + Logistic Regression pipeline yielded an accuracy of 96.8% and F1-score of 0.967 with a good precision-recall curve for both fake and true news. Consistent performance was witnessed with test samples with few misclassifications as evident from the confusion matrix.

This shows that even classic ML approaches, with good tuning and in conjunction with good preprocessing, can yield competitive performance in fake news detection. The pipeline is fast, lightweight, and good for practical use on low-resource systems.

6.References

Research paper : **Strengthening Fake News Detection : Leveraging SVM and Sophisticated Text Vectorization Techniques. Defying BERT? https://arxiv.org/abs/2411.12703**

Google Colab: https://colab.research.google.com/drive/12rUPne31p1qAHHXjIernSy_LB0J-JAG6#scrollTo=Gs7XTFalV3ZB

- Collected news articles and tweets related to the topic