



# Fake News Classification Using Machine Learning and Deep Learning Technique

Kumar Somendr<sup>1</sup>, Adnan Mahmood<sup>2</sup>

<sup>1, 2</sup> Department of Computer Science & Engineering, BIT Mesra, Patna Campus, Bihar, India.

## How to cite this paper:

Kumar Somendr<sup>1</sup>, Adnan Mahmood<sup>2</sup>, "Fake News Classification Using Machine Learning and Deep Learning Technique", IJIRE-V7I2-326-335.



Copyright © 2026  
by author(s) and  
Fifth Dimension  
Research

Publication. This work is licensed under the  
Creative Commons Attribution International  
License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>

**Abstract:** This project focuses on detecting fake news using a hybrid approach combining machine learning and deep learning techniques. The dataset includes real and fake news articles that are cleaned and preprocessed to improve quality. For feature extraction, TF-IDF is used to convert text into numerical form. Machine learning models such as Logistic Regression, Support Vector Machine, and Naive Bayes are trained and evaluated. Transformer-based models like BERT and RoBERTa are also implemented. Model performance is measured using accuracy, precision, recall, and F1-score, along with confusion matrix and ROC analysis. A majority voting technique is applied to improve final predictions. Results show that deep learning models provide better accuracy and reliability.

**Key Words:** Fake News Detection, Machine Learning, Deep Learning TF-IDF, BERT RoBERTa.

## I. INTRODUCTION

Fake news has become a major problem in today's digital world. With the rapid growth of the internet and social media platforms, information spreads very quickly among people. While this has many benefits, it also makes it easy for false or misleading information to spread. Fake news refers to incorrect or misleading information that is presented as real news. It is often created to influence people's opinions, create confusion, or gain attention.

In recent years, the spread of fake news has increased significantly due to platforms like social media, blogs, and online news websites. Anyone can create and share content without proper verification. As a result, it becomes difficult for users to identify whether the information they are reading is true or false. Fake news can have serious effects on society, including spreading misinformation, creating panic, and influencing political decisions.

Traditional methods of detecting fake news mainly relied on manual verification by journalists and fact-checkers. These methods include checking sources, verifying facts, and cross-referencing information with trusted sources. However, with the huge amount of data generated every day, manual verification is not efficient and cannot handle the speed at which fake news spreads. Therefore, there is a need for automated systems that can detect fake news quickly and accurately.

Machine learning and artificial intelligence provide effective solutions for this problem. These technologies can analyze large amounts of data and identify patterns that help in detecting fake news. In this project, different machine learning and deep learning models are used to classify news as real or fake. Traditional models such as Logistic Regression, Naive Bayes, and Support Vector Machine (SVM) are used along with advanced models like BERT and RoBERTa.

The process of fake news detection involves several steps. First, the dataset is collected and preprocessed. The text data is cleaned by removing unnecessary characters and converting it into a standard format. Then, feature extraction techniques such as TF-IDF are used to convert text into numerical form. After that, different models are trained on the data to learn patterns that distinguish fake news from real news.

The performance of these models is evaluated using various metrics such as accuracy, precision, recall, and F1-score. Confusion matrix, ROC curve, and learning curve are also used to analyze model performance. These evaluation methods help in comparing different models and selecting the best one.

The main goal of this project is to develop an efficient system that can automatically detect fake news with high accuracy. This system can be useful in real-world applications such as social media platforms, news websites, and online content verification tools. By using this system, users can easily identify whether a news article is real or fake. In conclusion, fake news detection is an important area of research in the field of machine learning and natural language processing. With the help of advanced models and techniques, it is possible to reduce the spread of fake news and create a more reliable information environment.

## Overview of Fake news

Fake news refers to false or misleading information that is presented as real news. It is often created to mislead people, gain attention, or influence opinions. The spread of fake news can create confusion, fear, and misunderstanding among people. It can also affect important areas like politics, health, and society. Therefore, detecting fake news has become very important.

To solve this problem, technologies like machine learning and transformer based models are used. These systems analyze the content of news and help in identifying whether it is real or fake.

### **II.OBJECTIVE**

The objective of this project is to examine the problems which is related with the spread of fake news . We will be working on fake news data set in which we will apply machine learning algorithm and transformer based model to train and test data to find which news is the real news or which one is the fake news . As the fake news is a problem that effects society and people perceptions on different topics . By using machines learning and transformer models , the problem can be solved as we able to find patterns from the data to maximize well defined objectives . so , our focus is to find fake news using both machine learning and transformer based models.

### **Traditional Approaches to News Verification**

Traditional approaches to news verification involve human-centered techniques that focus on assessing the accuracy and credibility of information without relying heavily on automated systems like machine learning. These methods have been used by journalists, fact-checkers, educators, and the public to verify the truthfulness of news. Although they can be time-consuming and less scalable, they offer transparency, human judgment, and contextual understanding, which are crucial in complex or sensitive situations.

#### **Here are the main traditional approaches:**

##### **1. Manual Fact-Checking**

In this method, people check if a news story is true or false by comparing it to trusted sources like government records or official websites. This is very accurate, but also slow.

##### **2. Crowdsourcing**

Here, many people on the internet help check news. They report, flag, or vote on posts they think are fake. It's fast and covers more posts, but not always right.

##### **3. Source Verification**

This means checking where the news came from. Trusted sites like BBC or The Hindu follow rules and correct mistakes

### **Machine Learning in Fake News Detection**

Scalable and precise detection techniques are desperately needed in light of the proliferation of bogus news on digital platforms. Although manual methods can be useful in certain situations, they are insufficient and time-consuming to manage the massive and rapidly changing streams of data found online. By identifying patterns in data and categorizing material appropriately, machine learning (ML) provides a strong and automated method for detecting fake news.

### **Data Collection and Pre-processing**

The first step in applying machine learning to detect fake news is gathering labeled datasets that contain instances of both true and fake news stories. These datasets are necessary for training machine learning model . Natural Language Processing(NLP) approaches are used for preprocessing before the input is fed into the model. Tokenization, stop-word elimination and text-to-numerical conversion using vectorization methods such as Term Frequency-Inverse Document Frequency (TF-IDF)

### **Machine Learning Algorithms**

For this objective, several machine learning techniques are used. Because of their effectiveness and interpretability, traditional classifiers like Support Vector Machines , Naive Bayes and Logistic Regression are frequently utilized. These models examine linguistic and statistical characteristics, including sentence structure, sentiment polarity, or word frequency.

### **Summary**

Fake news can be solved in a dynamic and scalable way with machine learning. Although not perfect, machine learning (ML) systems present a promising way to reduce misinformation and maintain the quality of online information, particularly when paired with human verification and conventional techniques.

### **Deep Learning Approaches in Fake News Detection**

Deep learning approaches use advanced neural network models to automatically detect fake news. These models learn patterns from large amounts of data and can understand the meaning of text better than traditional methods.

Models like BERT and Roberta are used in fake news detection. These models understand the context of words in both directions (left and right), which improves accuracy.

Although deep learning provides high accuracy, it requires a large amount of data. Training these models can be time-consuming, but they are very effective in detecting fake news in real-world applications.

### III. DATA COLLECTION

For this project, I used two public datasets: Fake.csv and True.csv. One had fake news articles, and the other had real ones

- 0 for fake news
- 1 for true news

This made it easy to train models.

Then, I merged both datasets into one file, shuffled the data to remove any pattern, and got it ready for cleaning and training. This careful setup helped build a clean and fair dataset for the project.

#### Dataset Description

Two distinct CSV files, false.csv and True.csv, which are both publicly accessible and frequently used in studies on false news detection, make up the dataset used in this project. News items from a variety of sources are included in these files, which are pre-labeled as "false" or "true" for binary classification, making them appropriate for supervised learning techniques

##### 1. Fake.csv

This file contains fabricated news articles that are either misleading, intentionally false, or lack factual basis

##### 2. True.csv

This file includes genuine news articles published by reputable news sources.

For the purpose of model training, only the text attribute was used, as it holds the complete article body required for content analysis and classification. A new column called class was added to both datasets:

1. Fake articles were labeled as 0
2. True articles were labeled as 1

In total, the combined dataset consists of thousands of labeled news articles, providing a balanced and diverse foundation for developing and evaluating machine learning models for fake news detection.

#### Data Preprocessing

Data preprocessing is the step where raw data is cleaned and prepared for the model. First, the dataset is loaded and unnecessary columns are removed. The text data is cleaned by removing special characters, numbers, and extra spaces, and converting all text to lowercase.

Then, common words (stopwords) are removed to keep only useful information. After that, the text is converted into numbers using TF-IDF so the model can understand it. Finally, the data is divided into training and testing sets to check the model's performance.

#### Data Labeling and Merging

To create an effective machine learning model for identifying fake news, a well labeled and cohesive dataset is essential. In this, two different datasets True.csv and Fake.csv were used. News articles are categorized according to their veracity in each dataset. To train a machine learning model, the data needs to be merged into a single dataset for consistent processing and have unambiguous class labels.

#### Data Labeling

Adding a new column named class to each dataset was the first step. While articles in the True.csv dataset were labeled with 1, signifying genuine or authentic news, those in the fraudulent.csv dataset were labeled with 0, signifying fraudulent news. Training classification models that can distinguish between the two groups requires this binary labeling.

#### Data Merging

After labeling, the two datasets were merged using pd.concat(). This created a single dataset combining both fake and true news. The merged dataset was then shuffled randomly to avoid

#### Text Cleaning

Text cleaning is the process of making text data clear and useful for the model. In this step, special characters, punctuation, numbers, and extra spaces are removed. All text is converted to lowercase to maintain consistency.

Common words like "the", "is", and "and" are removed because they do not add much meaning. This helps the model focus only on important words and improves accuracy.

#### Feature Extraction (TF-IDF Vectorization)

Feature extraction is the process of converting text into numerical form so the model can understand it. TF-IDF (Term Frequency–Inverse Document Frequency) is used to give importance to words based on how often they appear in a document and how unique they are across all documents.

Common words get lower values, while important and rare words get higher values. This helps the model focus on meaningful words and improves prediction accuracy.

Converting the processed textual data into numerical features that machine learning models can use is a critical next step in developing a fake news detection system.

### IV. MODEL SELECTION AND TRAINING

Model selection and training is the step where different machine learning and deep learning models are chosen and trained on the data. In this project, models like Logistic Regression, Naive Bayes, Support Vector Machine (SVM), BERT, and RoBERTa are used.

Each model learns patterns from the training data to classify news as real or fake. Traditional models like Logistic Regression, Naive Bayes, and SVM work on numerical features, while advanced models like BERT and RoBERTa understand the context of text more effectively.

After training, all models are tested on unseen data, and their performance is compared to select the best model for prediction.

#### Logistic Regression

Logistic Regression is a simple machine learning model used for classification. It predicts whether news is real or fake by calculating probability. It works well for basic text classification tasks.

#### Naive Bayes

Naive Bayes is a fast and simple algorithm based on probability. It assumes that all words are independent of each other. It is very useful for text data and works well for fake news detection.

#### Support Vector Machine (SVM)

SVM is a powerful model that separates data into different classes using a boundary line (hyperplane). It tries to find the best boundary to correctly classify news as real or fake.

#### BERT

BERT is a deep learning model used for understanding text. It reads words in both directions (left and right) to understand the full meaning of a sentence. This helps in detecting fake news more accurately.

#### RoBERTa

RoBERTa is an improved version of BERT. It is trained on more data and performs better in understanding text. It gives higher accuracy in fake news detection tasks.

#### Evaluation Metrics

Evaluation metrics are used to measure how well the model is performing. Accuracy shows how many predictions are correct out of total predictions. Precision tells how many predicted fake news items are actually fake, while recall shows how many actual fake news items are correctly identified.

The F1-score is a combination of precision and recall, giving a balanced result. A confusion matrix is also used to show correct and incorrect predictions. These metrics help in comparing different models and selecting the best one.

### V. PERFORMANCE OF MODELS

The performance of different models is evaluated using the test dataset. Models like Logistic Regression, Naive Bayes, SVM, BERT, and RoBERTa are trained and tested to classify news as real or fake.

Each model gives different accuracy and results based on how well it understands the data. Traditional models work faster, while deep learning models like BERT and RoBERTa provide better accuracy because they understand the context of the text.

#### Logistic Regression Results

The Logistic Regression model achieved an accuracy score of approximately 97%

```

=====
Logistic Regression Results:
=====
Accuracy : 0.9765
Precision: 0.9691
Recall   : 0.983
F1 Score : 0.976

Classification Report:
precision    recall  f1-score   support

0           0.98    0.97    0.98     2055
1           0.97    0.98    0.98     1945

 accuracy
macro avg   0.98    0.98    0.98     4000
weighted avg 0.98    0.98    0.98     4000
    
```

*SVM Results*

The svm model achieved an accuracy score of approximately 98%

```

=====
SVM Results:
=====
Accuracy : 0.9888
Precision: 0.9837
Recall   : 0.9933
F1 Score : 0.9885

Classification Report:
precision    recall  f1-score   support

0           0.99    0.98    0.99     2055
1           0.98    0.99    0.99     1945

 accuracy
macro avg   0.99    0.99    0.99     4000
weighted avg 0.99    0.99    0.99     4000
    
```

*Naive Bayes Results*

The Naive Bayes model achieved an accuracy score of approximately 93%

```

=====
Naive Bayes Results:
=====
Accuracy : 0.9307
Precision: 0.9216
Recall   : 0.9373
F1 Score : 0.9294

Classification Report:
precision    recall  f1-score   support

0           0.94    0.92    0.93     2055
1           0.92    0.94    0.93     1945

 accuracy
macro avg   0.93    0.93    0.93     4000
weighted avg 0.93    0.93    0.93     4000
    
```

*BERT Results*

The BERT model achieved an accuracy score of approximately 91%

Epoch	Training Loss	Validation Loss	Accuracy	Precision	Recall	F1
1	0.638072	0.536031	0.818750	0.826846	0.792793	0.809461
2	0.504861	0.448371	0.869375	0.825688	0.926641	0.873257
3	0.466898	0.406423	0.855625	0.790426	0.956242	0.865463
4	0.406109	0.364596	0.876875	0.819383	0.957529	0.883086
5	0.372354	0.328909	0.896875	0.871359	0.924067	0.896939
6	0.365916	0.317698	0.895000	0.848000	0.954955	0.898305
7	0.343722	0.298671	0.904375	0.872315	0.940798	0.905263
8	0.327391	0.288374	0.904375	0.871429	0.942085	0.905380
9	0.323919	0.285515	0.898750	0.854671	0.953668	0.901460
10	0.321608	0.274904	0.908125	0.873223	0.948520	0.909315
11	0.302902	0.269427	0.910625	0.878282	0.947233	0.911455
12	0.308077	0.267321	0.910625	0.876485	0.949807	0.911674
13	0.298717	0.265554	0.910000	0.874556	0.951094	0.911221
14	0.299100	0.264237	0.910000	0.874556	0.951094	0.911221
15	0.296325	0.263915	0.910000	0.874556	0.951094	0.911221

*Roberta Results*

The Roberta model achieved an accuracy score of approximately 95%

Epoch	Training Loss	Validation Loss	Accuracy	Precision	Recall	F1
1	0.645780	0.578417	0.824167	0.735257	1.000000	0.847433
2	0.542835	0.442285	0.902500	0.833570	1.000000	0.909232
3	0.427243	0.350369	0.894167	0.821879	1.000000	0.902232
4	0.379619	0.278624	0.920000	0.859238	1.000000	0.924290
5	0.300606	0.221658	0.944167	0.897397	1.000000	0.945924
6	0.305324	0.210018	0.930000	0.874627	1.000000	0.933121
7	0.285646	0.184729	0.940833	0.891933	1.000000	0.942880
8	0.258869	0.158596	0.954167	0.914197	1.000000	0.955175
9	0.250555	0.155473	0.950833	0.908527	1.000000	0.952071
10	0.267479	0.134368	0.960833	0.925750	1.000000	0.961444
11	0.291589	0.129646	0.960833	0.925750	1.000000	0.961444
12	0.261965	0.119604	0.965833	0.934609	1.000000	0.966200
13	0.282175	0.128531	0.960000	0.924290	1.000000	0.960656
14	0.286238	0.129486	0.959167	0.922835	1.000000	0.959869
15	0.261574	0.133350	0.956667	0.918495	1.000000	0.957516

*Confusion Matrix*

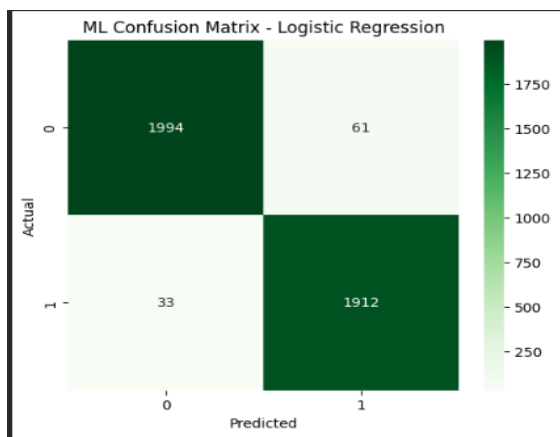
A confusion matrix is used to evaluate how well a model is performing. It shows the number of correct and incorrect predictions made by the model.

It has four parts:

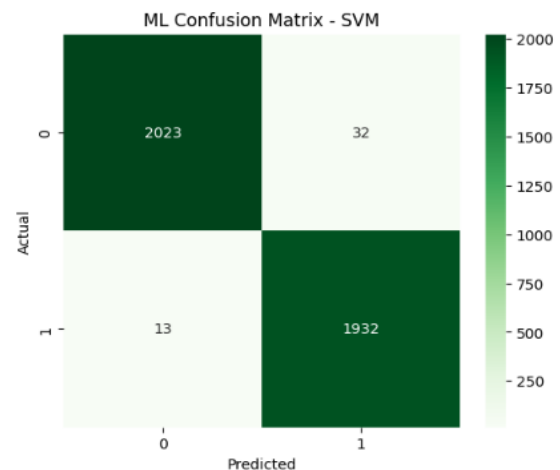
- **True Positive (TP):** Fake news correctly predicted as fake
- **True Negative (TN):** Real news correctly predicted as real
- **False Positive (FP):** Real news wrongly predicted as fake
- **False Negative (FN):** Fake news wrongly predicted as real

This matrix helps in understanding where the model is making mistakes and improves model performance.

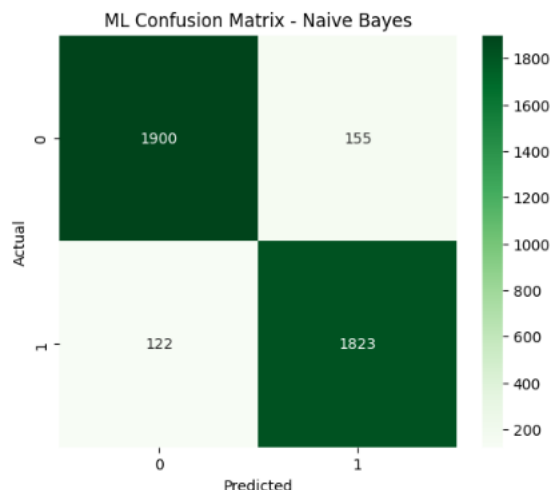
**Logistic Regression confusion matrix:**



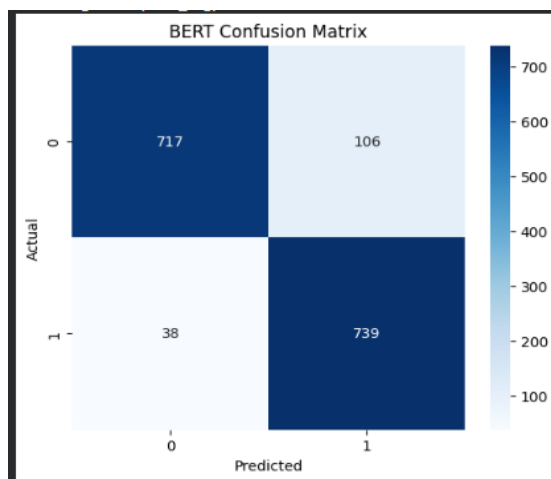
**Svm confusion matrix:**



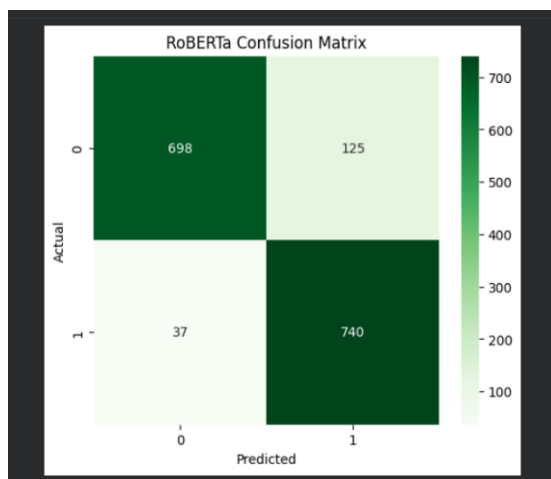
**Naive Bayes confusion matrix:**



**Bert confusion matrix:**



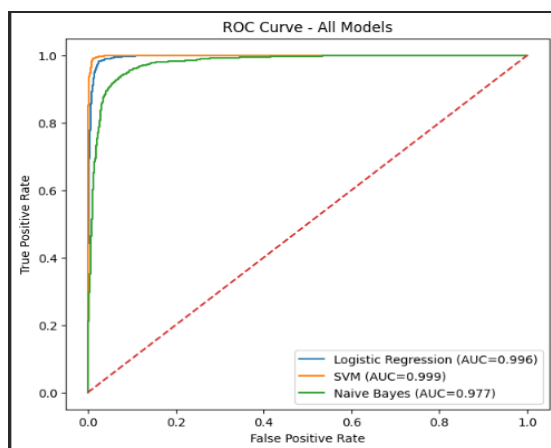
**Roberta confusion matrix:**



**ROC Curve Analysis**

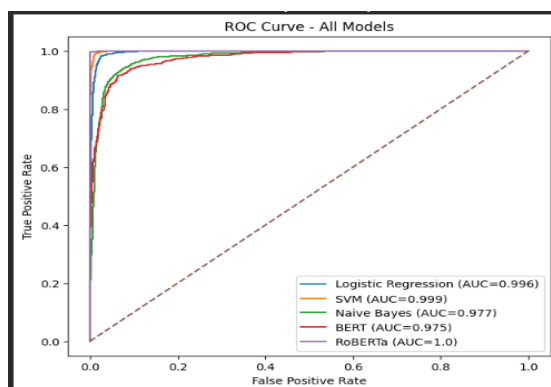
ROC (Receiver Operating Characteristic) curve is used to evaluate the performance of classification models. It shows the relationship between True Positive Rate (TPR) and False Positive Rate (FPR). A model with a curve closer to the top-left corner performs better. The Area Under the Curve (AUC) value indicates how well the model can distinguish between real and fake news. Higher AUC means better performance.

**ROC Curve for Machine Learning Models:**



This graph shows the ROC curves for Logistic Regression, Naive Bayes, and SVM models. All models perform well, but the model with the highest curve has better classification ability. This comparison helps in understanding which traditional model performs best.

**ROC Curve for All Models:**

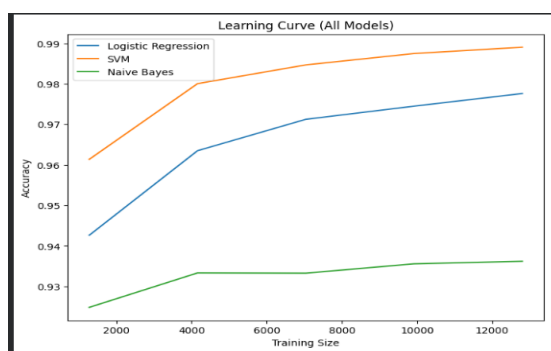


This graph shows the ROC curves for all models including Logistic Regression, Naive Bayes, SVM, BERT, and RoBERTa. It can be observed that deep learning models like BERT and RoBERTa perform better due to their ability to understand text context more effectively.

**Learning Curve Analysis**

Learning curve is used to understand how well a model learns from the training data. It shows the relationship between training size and model performance. Two curves are plotted: training score and testing score. If both curves are close, the model performs well. A large gap may indicate overfitting or under fitting.

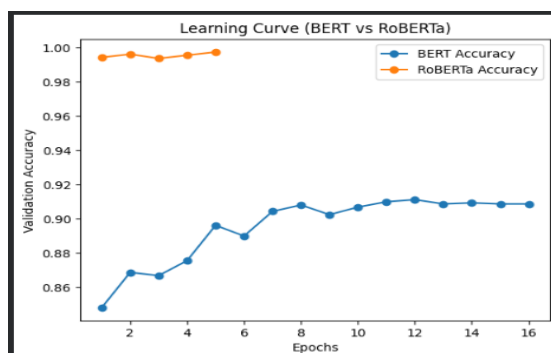
**Learning Curve for Machine Learning Models:**



## Fake News Classification Using Machine Learning and Deep Learning Technique

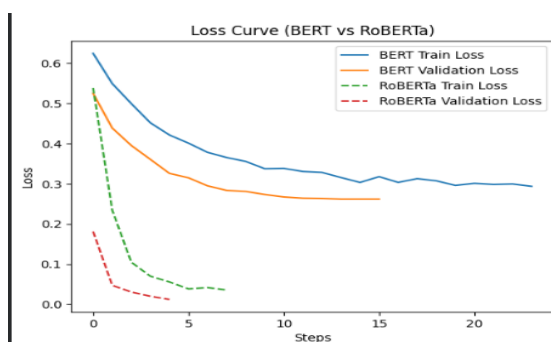
This graph shows the learning curves of Logistic Regression, Naive Bayes, and SVM. The performance improves as the training data increases. The small gap between training and testing curves indicates good generalization.

### Learning Curve for Deep Learning Models:



This graph represents the learning rate or accuracy curve of BERT and RoBERTa models. It shows how model performance improves during training epochs.

### Loss Curve for Deep Learning Models:



This graph shows the loss curve of BERT and RoBERTa. The decreasing loss value indicates that the model is learning effectively and improving over time.

## VLMANUAL TESTING AND PREDICTION

Manual testing is performed to check how the model works on user input. In this, the user enters a news text and the system processes it using the trained models.

The input text is first cleaned and converted into numerical form using the same preprocessing steps. Then, all models make predictions on the input data. The final result is decided based on the model outputs.

### Model predictions

```
==== MODEL PREDICTIONS ====
Logistic Regression : Fake News
SVM                  : Fake News
Naive Bayes          : Real News
BERT                  : Real News
RoBERTa              : Real News

==== FINAL RESULT ====
Final Prediction : Real News
```

## VII.CONCLUSION

Many people consume news from social media instead of traditional news media. However, social media has also been used to spread fake news, which has negative impacts on individual people and society. In this paper, an innovative model for fake news detection using machine learning algorithms has been presented. The model takes news from users input and based on classification algorithms it predicts if news is fake or true

### Future scope

In the future, the fake news detection system can be improved by using larger and more diverse datasets to increase accuracy. More advanced deep learning models can also be added to enhance performance.

The system can be extended to detect fake news in multiple languages and from different sources like social media, images, and videos. Real-time detection can also be implemented so that users can verify news instantly.

In addition, the model can be integrated into web applications or browser extensions to help users easily identify fake news while browsing.

### References

1. Guan, R., Zhang, H., Liang, Y., Giunchiglia, F., Huang, L. and Feng, X., 2020. Deep feature-based text clustering and its explanation. *IEEE Transactions on Knowledge and Data Engineering*, 34(8), pp.3669-3680.
2. Bouveyron, C., Celeux, G., Murphy, T.B. and Raftery, A.E., 2019. *Model-based clustering and classification for data science: with applications in R (Vol. 50)*. Cambridge University Press.
3. Zong, C., Xia, R. and Zhang, J., 2021. *Text data mining (Vol. 711, p. 712)*. Singapore: Springer.
4. Kumar, S., & Singh, M. (2020). A Survey on Fake News Detection using Machine Learning Techniques. *International Journal of Advanced Research in Computer Science and Software Engineering*, 10(7), 1-5.
5. scikit-learn: Machine Learning in Python – <https://scikit-learn.org/>
6. Kaggle – <https://www.kaggle.com/>
7. True.csv and fake.csv : <https://www.kaggle.com/datasets/emineytm/fake-news-detection-datasets>
8. Pandas Documentation – <https://pandas.pydata.org/docs/>
9. I. Ahmad, M. Yousaf, S. Yousaf, et al., “Fake news detection using machine learning ensemble methods,” *Complexity*, pp. 1-11, 2020
10. W. He, Y. He, B. Li, et al., “A naïve-Bayes-based fault diagnosis approach for analog circuit by using image-oriented feature extraction and selection technique,” *IEEE Access*, vol. 8, pp. 5065- 5079, 2020.
11. Q. Xue, Y. Zhu, and J. Wang, “Joint distribution estimation and naïve bayes classification under local differential privacy,” *IEEE Transactions on Emerging Topics in Computing*, vol. 9, pp. 2053- 2063, 2021.
12. H. A. Maddah, “Decision trees based performance analysis for influence of sensitizers characteristics in dye-sensitized solar cells,” *Journal of Advances in Information Technology*, vol. 13, pp. 271- 276, 2022.
13. I. D. Mienye, Y. Sun, and Z. Wang, “Prediction performance of improved decision tree-based algorithms: A review,” *Procedia Manufacturing*, vol. 35, pp. 698-703, 2019.
14. J. A. C. Moreano and N. B. L. S. Palomino, “Global facial recognition using gabor wavelet, support vector machines and 3D face models,” *Journal of Advances in Information Technology*, vol. 11, pp. 143-148, 2020.
15. A. B. Gumelar, A. Yogatama, D. P. Adi, et al., “Forward feature selection for toxic speech classification using support vector machine and random forest,” *International Journal of Artificial Intelligence*, vol. 11, pp. 717-726, 2022.
16. J. Cervantes, F. García-Lamont, L. Rodríguez, et al., “A comprehensive survey on support vector machine classification: Applications, challenges and trends,” *Neurocomputing*, vol. 408, pp. 189-215, 2020.
17. I. Benchaji, S. Douzi, and B. E. Ouahidi, “Credit card fraud detection model based on LSTM recurrent neural networks,” *Journal of Advances in Information Technology*, vol. 12, pp. 113- 118, 2021.
18. N. Yadav and A. K. Singh, “Bi-directional encoder representation of transformer model for sequential music recommender system,” in *Proc. Forum for Information Retrieval Evaluation*, 2020, pp. 49- 53.
19. Brown, T., et al. (2020). *Language Models are Few-Shot Learners*.
20. Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*.