Fake News Detection

Mohamed Awad EL-Tahan**: N-grams Feature**

Supervisor: Dr. Wesam Ahmed

Group Number: 3

Submission date: May 12th, 2025

Code Implementation

Preprocessing:

1. All text was converted to lowercase.

2. Remove punctuation and numbers.

3. Remove stop words using NLTK library.

4. Each word was lemmatized using NLTK library (Lemmatization).

5. The text was split into individual words using NLTK library

   (tokens).

Visualization:

Figure 2.1: Word cloud

Text Transformation:

We used three transformation techniques: -

1. Splitting Text into Tokens.

2. .

3.

Model	Training Accuracy	Testing Accuracy	Accuracy	Recall	Precision	F1-measure	AUC Value
Logistic	100%	92.49%		96.4 %	91.11%	93.68%	98%
Nave Bayes	99.7%	93.37%		92.66%	95.71%	94.16%	98%
SVM	93.45%	90.65%		91.07%	92.61%	91.84%	97%
Random Forest	100%	88.78%		97.58%	85.14%	90.94%	97%
LSTM	99.27%	93.41%		95.5%	93.24%	94.36%	97%
CNN	99.18%	93.09%		92.8%	95.11%	93.94%	86%
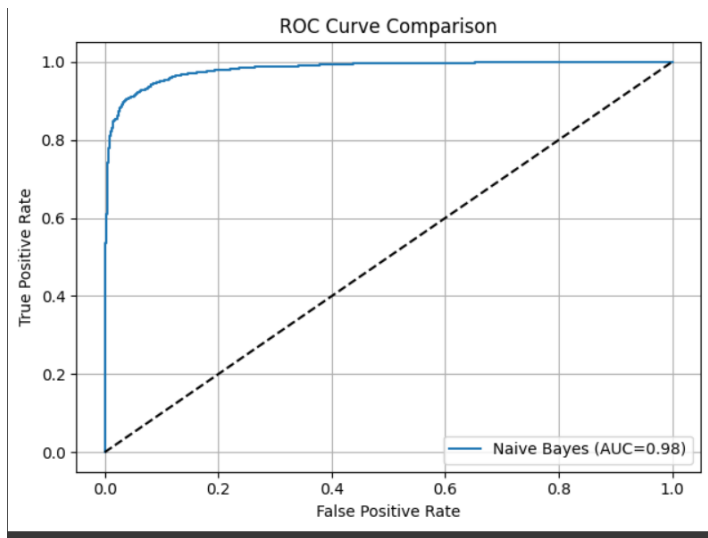
Fig 1: Logistic Regression Roc curve
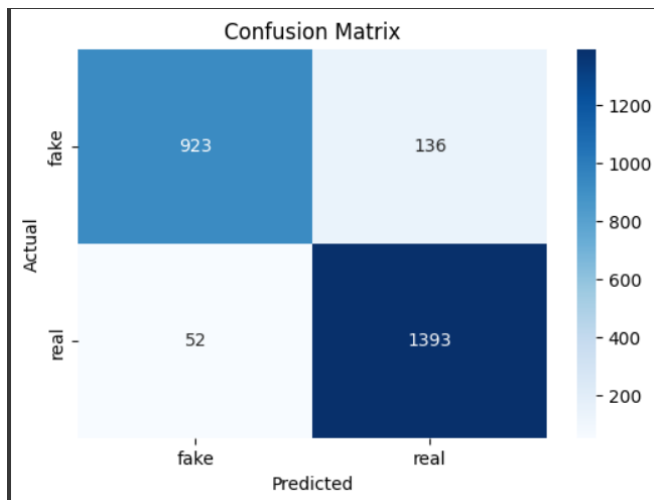
Fig 2: Logistic Regression Confusion Matrix

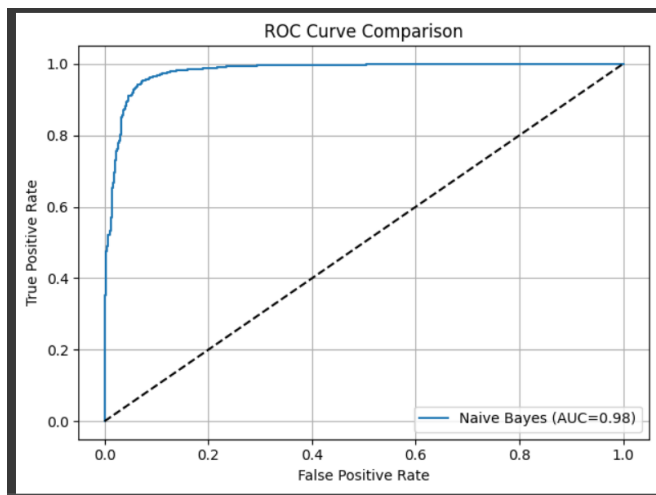Fig 3: Nave Bayes Roc curve

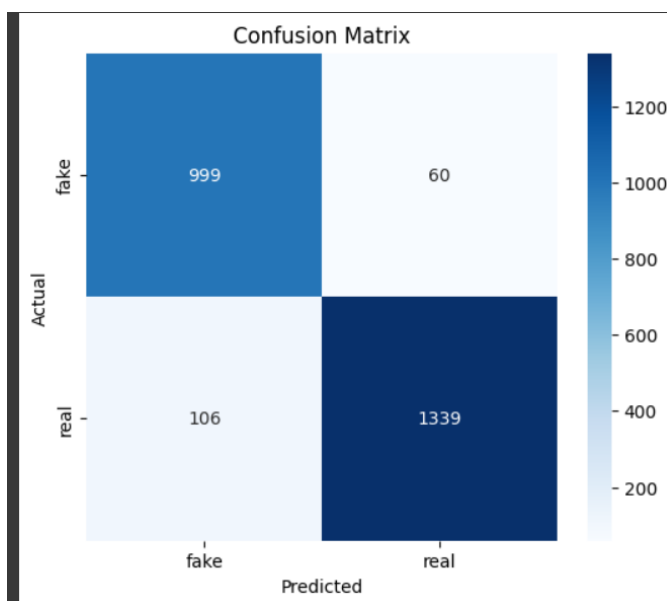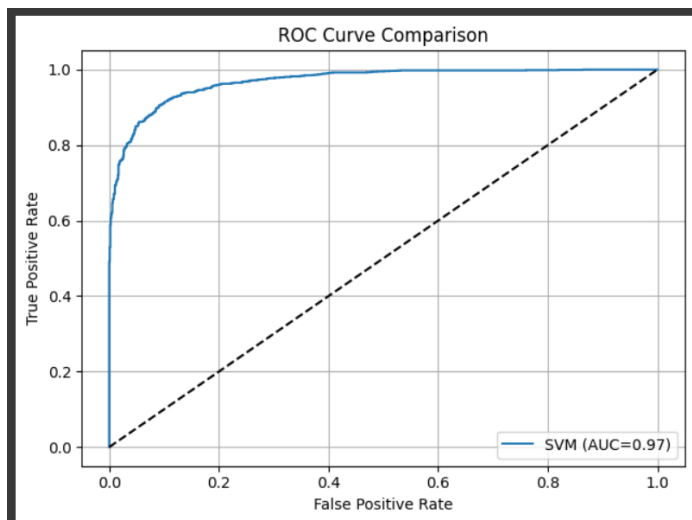Fig 4: Nave Bayes Confusion Matrix
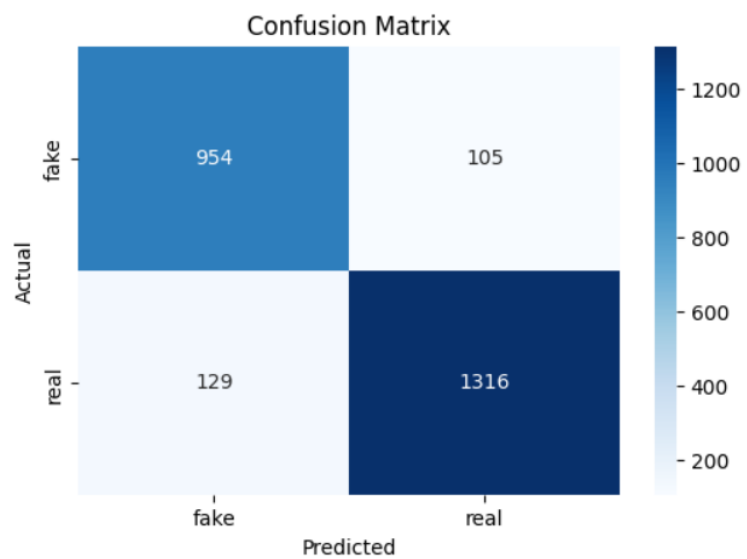
Fig 5: SVM Roc curve

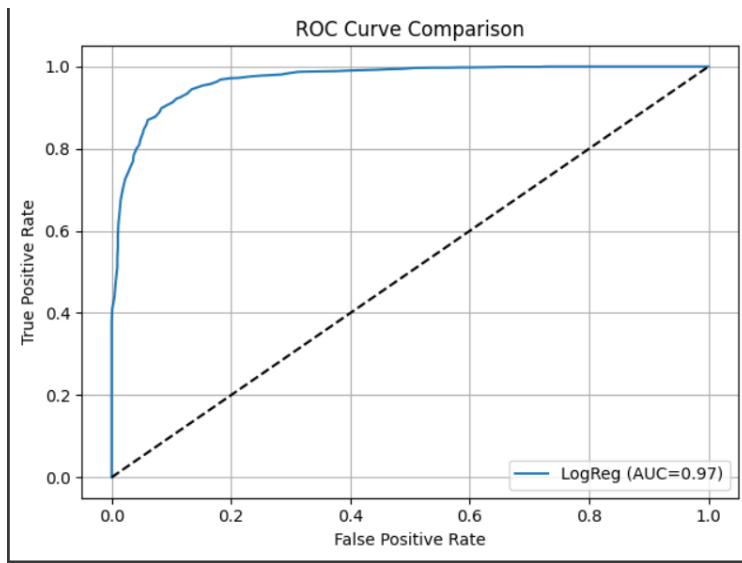Fig 6: SVM Confusion Matrix

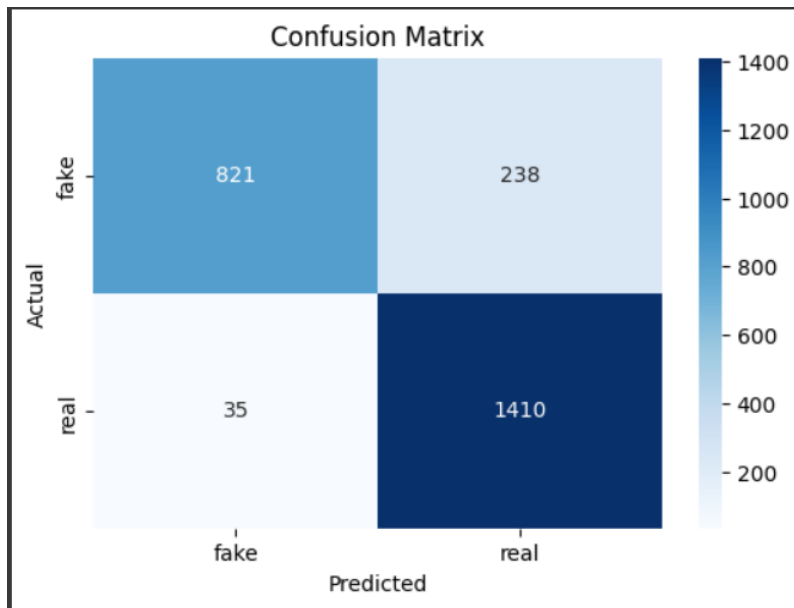Fig 7: Random Forest Roc curve
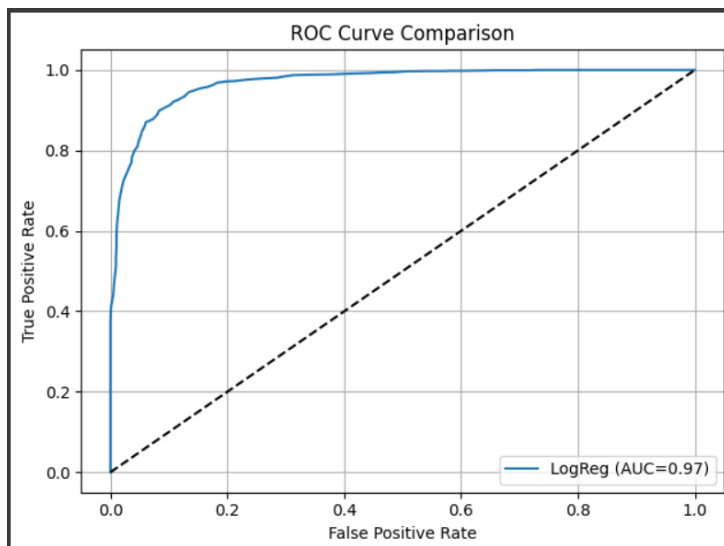

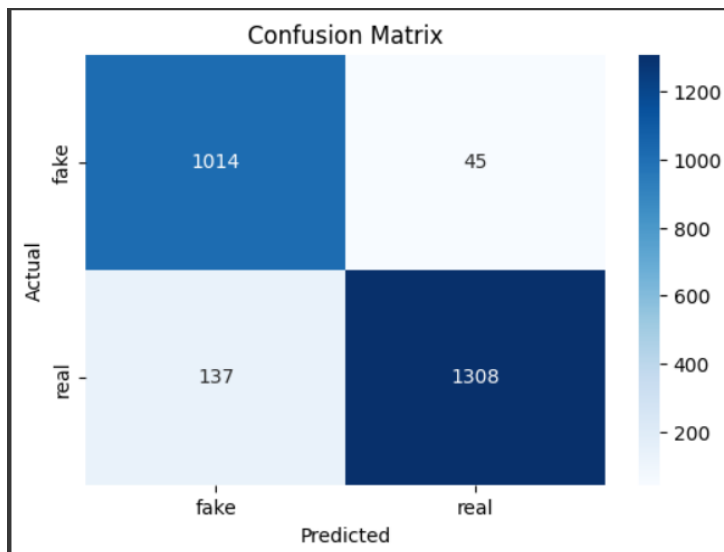Fig 8: Random Forest Confusion Matrix
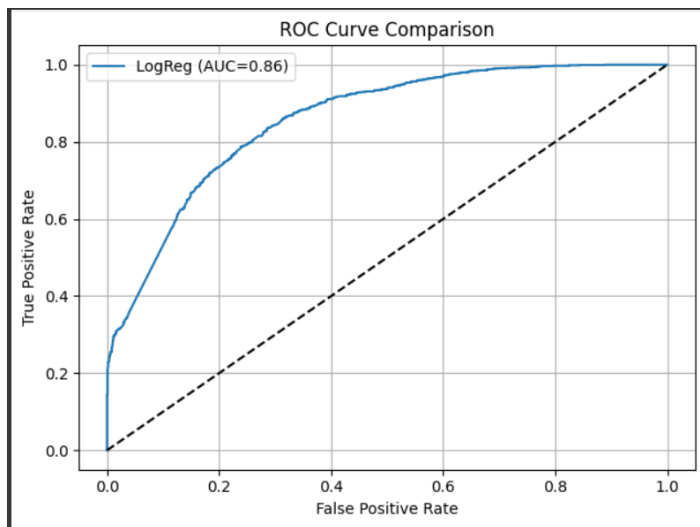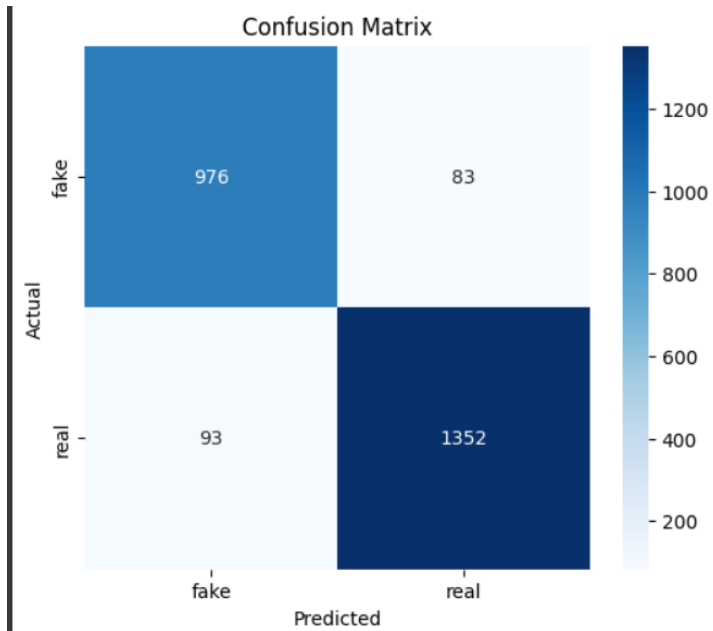
Fig 9: LSTM Roc Curve

Fig 10: LSTM Confusion Matrix

Fig 11: CNN Roc Curve

Fig 12: CNN Confusion Matrix