Kurdish Studies
July 2024
Volume: 12, No: 5, pp 1008 -1015
ISSN: 2051-4883 (Print) | ISSN 2051-4891 (Online)
www.KurdishStudies.net

DOI: 10.53555/ks.v12i5.3390

Comparative Study of Fake News Detection using Navie Bayes and Logistic Model

Naveed Sheikh[1*], Sajida Parveen[1], Abdul Rehman[1], Asma Naeem[1], Misbah Anjum[1], Komal Shahid[1], Muhammad Yasin[2]

[1]Department of Mathematics, University of Balochistan, Quetta, Pakistan.
[2]Department of Mathematics, Balochistan University of Information Technology, Engineering and Management Sciences, Quetta, Pakistan

***Corresponding Author: Naveed Sheikh**
[1*]Department of Mathematics, University of Balochistan, Quetta, Pakistan. Tel: +923358396878, Email: naveed.maths@um.uob.edu.pk

Abstract
In this research, machine learning methods are used to automatically recognize fake news which is currently a major threat on the web regarding acquired electronically information validity and accuracy. After that, calculate the performance of both the Logistic Regression (LR) model and the Navie Bayes (NB) classifier to avoid fake news. Comparative analysis suggests that Logistic Regression (LR) model delivers higher accuracy than Navie Bayes (NB) classifier in fake news detection. The performance of both classifiers is calculated by confusion matrix, (ROC-AUC) Curve and metric evaluation report for which cross-validation technique used. Accuracy achieved by Logistic Regression (LR) Model is around 98%, overall average accuracy in correctly identifying a fake news compared to Naive Bayes (NB) classifier which could only achieve approximately 93 %, keys matrices such as recall, precision, recall, F1-Score, F2-Score computed on both the models gave values closer to 0.98 at Logistic end better than Navie Bayes (NB) best obtained value of corresponding metrics i.e., 0.93. In addition, the K-fold cross validation is also applied to illustrate our generalized predictions for unknown data. Both models showed reasonable statics in performance. A Logistic Regression (LR) classifier scored better in accuracy than a Navie Bayes (NB) classifier and had a mean prediction of 96% vs 93% for the Navie Bayes (NB) model. Other computed metrics backed this up as well. The large performance gap calculated demonstrates the great utility of logistic model over Naive Bayes (NB) where this can lead to handle the problem under misinformation. For fake news given different thresholds, future research should focus on how well the applied probabilistic calibration responds to real-world application scenarios (ROC-AUC).

Introduction

A digital era has been created with the rise of the internet; every individual is hooked to the various types of web resources (Saeed et al. 2018) . Information travels faster to a larger group in seconds. In the past, news was read through newspapers and now we are used to daily updating apps like Facebook, Twitter, WhatsApp forwards and downloading news article online (Hussain et al. 2023). As such time was internet users follow the sports events online and in addition to this mobile devices were spread all over so it makes them more easy saving there time (Saleem et al. 2023). News on social media is a two-edged sword. Social media has made it quite easy for the general public to obtain news and information, so people have started consuming this as the primary channel in which they get their daily dose of what is happening around them (Sherazi et al. 2023).

Meanwhile, it encourages the propagation of fake news, which is distinguished by poor and low-quality information, having a detrimental influence on both society and individuals and contributing to a culture of distrust (Ahmad et al. 2020). Digital media, a great source of information, is routinely abused. The increasing risk of fake news not only adds to the complexity of the matter but also greatly affects public perceptions and attitudes to the reliability and credibility of internet resources, so detection of fake news has been a challenging task for researchers and has taken the attention of many of them as a prominent area of research (Shu et al. 2017). The process of distinguishing false news from a vast amount of information poses a unique combination of challenges and complexity (Allcott and Gentzkow 2017). Many approaches have been developed throughout time to address these types of difficulties, including Natural Language Processing (NLP), Text Classification, and Sentiment Analysis. Researchers conduct several research efforts in the identification of false news (Jwa et al. 2019). The spread and dependability of information on the internet is a major issue in modern society; false news and misinformation are on the rise, and they have a significant impact on political and social reality (Yang et al., n.d.). The word vector training Mode approach, which is based on deep learning techniques, was presented to investigate twitter syntax and produce a binary classifier. The method was tested on a dataset of genuine tweets collected over ten days. The results show that this approach outperforms previous text-based and non-text-based spam detection strategies. (Wu et al. 2017). The main characteristics underlying fake news shared from both traditional media and social media are used during the detection phase to filter out false contents from the data set. The mining perspective of the data set, which includes feature extraction and model construction, as well as evaluation metrics for fake news detection, is also discussed (Aldwairi and Alwahedi 2018). An automatic hoax detection system is employed to detect fake news. It is primarily based on machine learning, with context analysis and a social context model combined to develop an Machine Learning (ML) fake news detection method that was implemented within Facebook

Messenger and achieved an accuracy of 81.7%. (Della Vedova et al. 2018). Fraudulent information and disinformation are readily and quickly spread by unsupported sources, hence an automated control system for detecting fake cases in Twitter threads using Support Vector Machine (SVM) has been created. On each user's profile, a reputation-based method employing a Machine Learning (ML) algorithm has been utilized to construct sentiment score according to the user's history and information (Deokate 2019). A new model for real-time spam in Twitter has been proposed and applied to a sample date of 40,000 public tweets. For categorizing the tweets, the top 30 words are extracted to obtain the highest information gain and generated 91% accuracy while compared to real-time tweets using four distinct Machine Learning (ML) algorithms: Support Vector Machine (SVM), Artificial Neural Network (ANN), Random Forest (RF), and Gradient Boosting (GB) (Gupta et al., n.d.). Another approach, Tri-Relationship bogus news detection (TriFN), was utilized to detect fake news. It relies on interactions between publishers, articles, and the audience to detect misleading information and has shown to be a more accurate and dependable way than other methods of identifying fake news, particularly in real-world contexts (Shu, Wang, and Liu 2019).

In this paper, a comparative analysis of Logistic Regression (LR) and Navie bayes (NB) classifier has been employed to identify fake news. Performance of both classifiers are evaluated and analyzed using evaluated set of metrics such as including precision, recall, F1-Score, (ROC-AUC) score, and Matthews Correlation Coefficient (MCC), also using technique of k-fold cross validation.

Methodology

i. **Data Collection.** Data is collected from Kaggle's fake news dataset containing fake and real news, this data is downloaded and saved in CSV file.

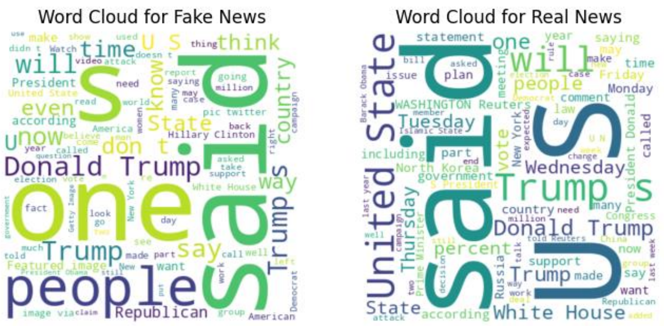

Figure 1. Word cloud for fake and real news.

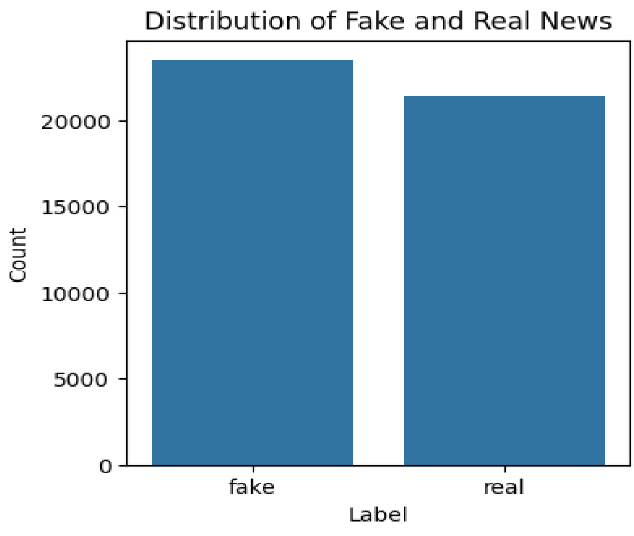

Figure 2. Count plot of fake and real news in data set.

Datasets for fake and real news, is loaded and labeled as "Fake or Real" word clouds are generated to highlight the most frequent and common words in fake and real news in the date set as given in figure 1. Fake and real news in the data set is plotted using count plot as shown in figure 2. Word count plot for real and fake pieces of news represent the most frequent or most occurring words in data set. This plot visualizes and highlights common themes or topics, language and pattern of contents lie in the real news and fake news articles.

ii. Data Preprocessing

Data preprocessing is a key step in data analysis and modeling because it prepares text data to be transformed and used in subsequent investigations. The primary goal of preprocessing is to clean up the data, remove noise and unnecessary characteristics, and extract relevant and meaningful features for analysis. The data preparation approach involves the following steps:

a) **Tokenization**

Tokenization is the process of dividing data into words known as tokens, with each token representing a single unit. Tokenization is accomplished by eliminating irrelevant white spaces, punctuation, and special instances and converting text data to lowercase for analysis.

b) **Stop word removal**

Words that do not impact meaning, such as "the" and "a," are removed to minimize noise and enhance model performance.

c) **Vectorization**

Vectorization is the process of converting text tokens into numerical vectors with high dimensions to make data easier to comprehend by machine learning algorithms.

d) **Feature extraction**

Extract primary and significant characteristics from text data and transform them into numerical form, which may be done using sentiment analysis and TF-IDF (Term Frequency - Inverse Document Frequency). The goal is to minimize the data dimensionality and provide a comprehensible pattern of text data.

e) **Normalization**

It is the scaling of data from 0 to 1 that enhances model convergence and performance.

f) **Transformation**

Data is transformed to make the text data acceptable for machine learning needs, hence increasing model reliability.

Model Selection

i. **Logistic Regression (LR) Model**

The algorithm was developed by statisticians to discuss population growth in ecology, but was later adopted by machine learning as a method to classify data, especially data that has the primary task of determining whether or not it belongs to a certain class. Logistic function is a supervised machine learning model also called sigmoid function. It determines whether a given instance belongs to a given class. Based on a set of independent variables, it estimates binary values like 0/1, yes/no, and true/false. The output values fall between 0 and 1. It is primarily used to categorize data into binary categories. The Logistic Regression (LR) model is formally expressed as, the logit odds of probability are modelled simply a linear function of the predictor feature values as:

$$logit(p) = ln\left(\frac{p}{1-p}\right) = a_0 \ + a_1X_1 \ a_2X_2 + \ldots + a_kX_k$$

ii. **Navie Bayes (NB) Classifier**

This is a supervised Machine Learning (ML) model based on probability theory, considered to be a fundamental concept in Machine Learning (ML). The assumptions are that features are independent of each other, mathematically expressed as:

$$P(c\backslash y) = \frac{P(y\backslash c) * P(c)}{P(y)}$$

This model calculates the probability $P(c\backslash y)$ of posterior class given that likelihood of input features $P(y\backslash c)$. $P(y)$ *and* $P(c)$ are marginal and prior probabilities of feature and class respectively. Although Navie Bayes (NB) is easily implemented, it is not suitable for data sets that contain complex relationships between features. It mainly deals with independent features, which is not the case in real-life data sets.

Results and Discussion

The main of study is to create a model based on machine learning techniques which can effectively work for detection of fake news in articles which is rising demand of today's world with the spreading misuse of social media in respective of misinformation. The model has been evaluated based on confusion matrix, classification report, receiver operating characteristic curve and area under the curve (ROC-AUC), evaluation metrics and cross validation will be discussed in next section and achieved accuracy of model about 98%.

i. **Confusion Matrix**

Confusion matrix visually display the performance of model on the test data and give the summary of prediction against actual values, it gives counts values of corrected and incorrected predictions made by model, significantly it gives insight how much model is confused while making predictions. The performance of Logistic Regression (LR) and Naive Bayes (NB) classifier in distinguishing fake news in context to respective confusion matrices.

Table 1. Confusion matrix for Logistic Regression (LR) model.

Total = 8980	**Fake news predicted**	**Real news predicted**
Fake news actual	(4182) TP	(65) FN
Real news actual	(79) FP	(4654) TN

Table 2. Confusion matrix for Navie bayes (NB) classifier.

Total = 8980	**Fake news predicted**	**Real news predicted**
Fake news actual	(3919) TP	(328) FN
Real news actual	(276) FP	(4457) TN

Confusion matrix for Logistic Regression (LR) model clearly indicates the correct classification of 4182 fake news pieces as fake and 4654 real news pieces as real out of 8980 news instances given in the data set, but the model misclassified 65 fake news as real and 79 real news pieces as fake leading to error in model as illustrated in table1.
Confusion matrix for Navie Bayes (NB) for the same data set estimated 3919 fake as fake instances and 4457 real as real news correctly, as listed in table 2. The above-described facts concluded that Logistic Regression (LR) model outperforms the Naive Bayes (NB) model in identifying fake news based on information displayed by confusion matrices.

Classification Report

A classification report is the summary of the overall performance of both classifiers Its calculations are based on the information obtained from confusion matrix i.e. the precision, recall, specificity, F1-Score, F2-Score, macro average and weighted average for given support (number of instances in each class) for fake and real news are computed. All metrics are calculated for 4733 instances of fake news articles and that of 4247 for real news articles as indicated in table 3.
The higher values of all metrics obtained in classification report is an indication of overall better performance of logistic classifier in the detection of fake and real news in the test data set and achieved 98% accuracy, phenomenally well in detecting both fake and real news, while Navie bayes classifier with very few errors in both cases of false positives and false negatives predictions, exhibited accuracy of 93%. Logistic Regression (LR) model achieved the higher values of accuracy, precision, recall, and F1-Score demonstrated overall optimization in performance. These findings suggest that Logistic Regression (LR) is more appropriate for the task of fake news detection.

Table 3. Classification report for Logistic Regression (LG) and Navie Bayes (NB) Classifier.

News type fake and real total instances = 8980	**Classification Report for Logistic Regression (LR) Model**						
	Precision	**Recall**	**F1-Score**	**F2-Score**	**Specificity**	**Support**	**Overall accuracy**
Fake	0.9862	0.9833	0.9847	0.9847	0.9833	4733	
Real	0.9814	0.9846	0.9830	0.9830	0.9846	4247	
Macro Average	0.9838	0.9840	0.98	0.9839	0.9839	8980	
Weighted average	0.9839	0.9839	0.98	0.9839	0.9839	8980	
	Classification Report for Navie Bayes Classifier						
News type fake and real total instances = 8980	**Precision**	**Recall**	**F1-Score**	**F2-Score**	**Specificity**	**Support**	**Overall accuracy**
Fake	0.93	0.94	0.94	0.93	0.94	4733	0.93 93%
Real	0.93	0.92	0.93	0.92	0.92	4247	
Macro average	0.93	0.93	0.93	0.92	0.93	8980	
Weighted average	0.93	0.93	0.93	0.92	0.93	8980	

Receiver Operating Characteristic Curve and Area under the curve (ROC-AUC)

The Receiver Operating Characteristic (ROC) and Area Under the Curve (AUC) are standard metrics are computed and plotted to evaluate the performance of our model for fake news detection, The (ROC) curve is obtained by plotting True Positive Rate (TPR) and the False Positive Rate (FPR) at different threshold set values as shown in figure 3 & 4.

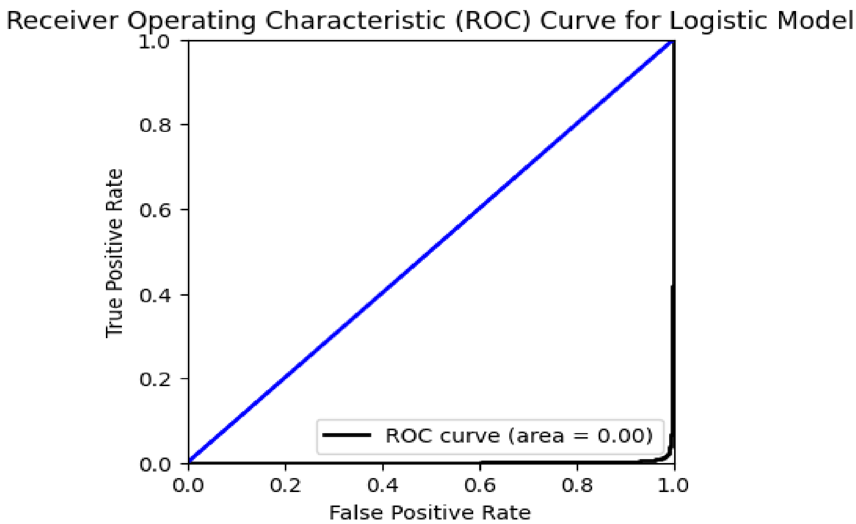

Figure 3. (ROC-AUC) plot for Logistic Regression (LR) classifier.

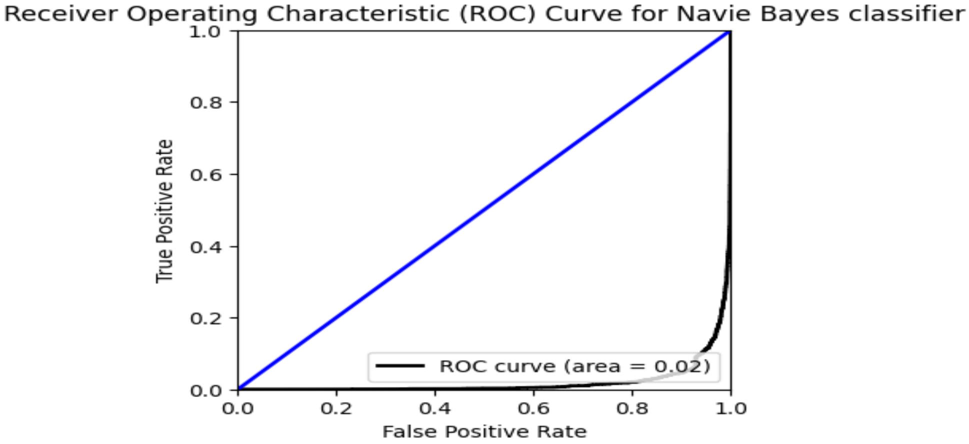

Figure 4. (ROC-AUC) plot for Navie Bayes (NB) classifier.

These values of (ROC-AUC) depict how well the classifier distinguish two classes of fake and real news. (ROC-AUC) is in the form of single value usually ranged from 0 to 1 describes overall performance and effectiveness of model. AUC = 0 & 0.02 for Logistic Regression (LR) and Navie Bayes (NB) classifier respectively which is indication of further amendment in the data set for future work. Both models are performing well for some specific threshold but not well for all values of threshold values for data set, which is improved by optimizing the data set. The Logistic Regression (LR) model attained higher values of (ROC-AUC) = 0.99 relative to that of 0.97 for Navie bayes (NB) classifier, indicating better performance in distinguishing between fake and real news.

Evaluation Metrics

Some other useful metrics such as Matthews Correlation Coefficient, Balanced accuracy, Cohen's Kappa Score are calculated as given in table 4. All these metrics take the values closer to 1, leads to better performance of model in detection of fake news.

Table 4. Calculation of Matthews Correlation Coefficient (MCC), Balance accuracy and Cohen's Kappa score.

Metrics/models	Matthews Correlation Coefficient (MCC)	Balanced Accuracy	Cohen's Kappa Score (kappa)
Logistic Regression (LR) model	0.9678	0.9840	0.96783
Navie Bayes (NB) classifier	0.8650	0.9322	0.8649

The Logistic Regression (LR) model outperforms the Naive Bayes (NB) model with higher balanced accuracy, indicating better performance on imbalanced data. The (MCC) value is higher for the Logistic Regression (LR) model, showing it has a stronger correlation between the actual and predicted classifications. The Logistic Regression (LR) model shows higher agreement between actual and predicted values. All the computed metrics are plotted against score represented, taking the score for each metric on y-axis, which ranges from 0 to 1, all values are getting closer to 1 as displayed in figure 5 & 6. It gives insight of different behavior of metrics briefly and overview of overall performance and effectiveness of model. The calculated metrics collectively provide a comprehensive examination of the model's effectiveness in detecting fake news.

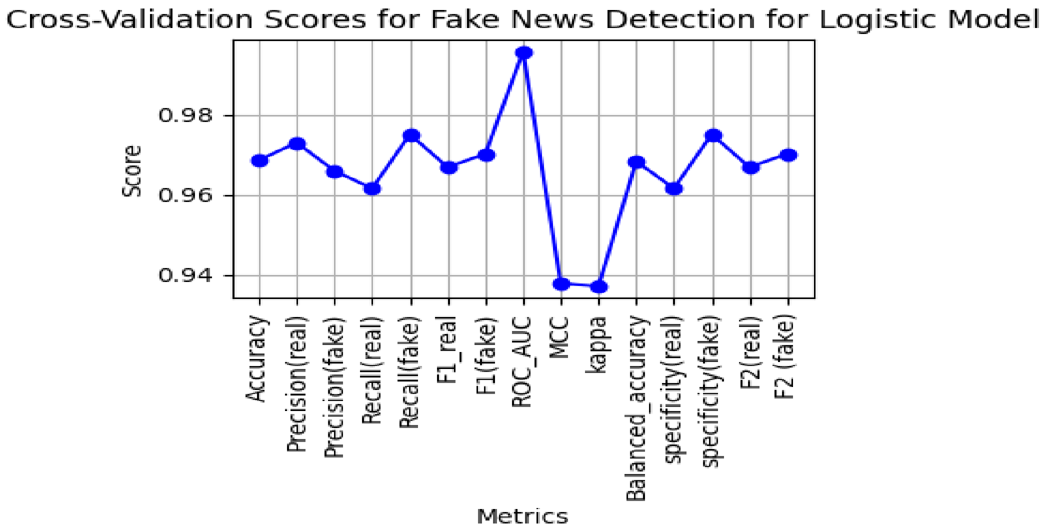

Figure 5. Evaluation metrics plot for Logistic Regression (LR) model.

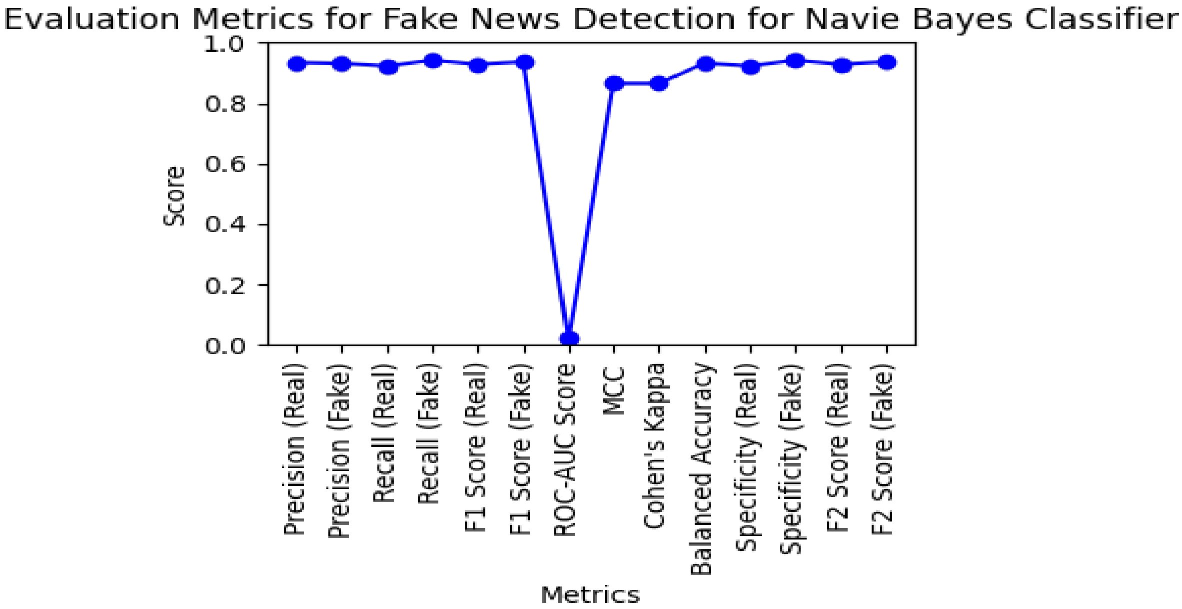

Figure 6. Plot of Evaluation metrics and score for Navie Bayes (NB) Classifier.

Cross Validation

Cross-validation which employed is a statistical approach to evaluate the overall performance and explain generalization ability of our fake news detection model. The results are listed in table 5.

Table 5. Cross validation of metrics for Logistic Regression (LR) and Navie bayes (NB) classifier.

Metrics	Cross validation values for Logistic Regression (LR) model	Cross validation values for Navie Bayes (NB) classifier
Accuracy	0.9687 (+/- 0.0071)	0.9345 (+/- 0.0032)
Precision (Real)	0.9730 (+/- 0.0173)	0.9328 (+/- 0.0045)
Precision (Fake)	0.9660 (+/- 0.0200)	0.9360 (+/- 0.0030)
Recall (Real)	0.9616 (+/- 0.0232)	0.9299 (+/- 0.0034)
Recall (Fake)	0.9751 (+/- 0.0167)	0.9387 (+/- 0.0043)
F1-Score (Real)	0.9669 (+/- 0.0078)	0.9314 (+/- 0.0032)
F1-Score (Fake)	0.9702 (+/- 0.0065)	0.9373 (+/- 0.0031)
ROC -AUC	0.9958 (+/- 0.0023)	0.9794 (+/- 0.0009)
(MCC)	0.9379 (+/- 0.0142)	0.8687 (+/- 0.0063)
Kappa Score	0.9372 (+/- 0.0143)	0.8687 (+/- 0.0063)
Balanced accuracy	0.9684 (+/- 0.0075)	0.9343 (+/- 0.0031)
Specificity (Real)	0.9616 (+/- 0.0232)	0.9299 (+/- 0.0034)
Specificity (Fake)	0.9751 (+/- 0.0167)	0.9387 (+/- 0.0043)
F2-Score (Real)	0.9669 (+/- 0.0078)	0.9314 (+/- 0.0032)
F2-Score (Fake)	0.9702 (+/- 0.0065)	0.9373 (+/- 0.0031)

The training subset of the data was divided into different subsets, and the validation subset was divided into smaller subsets. Repeating this process 1000 times provides more reliable estimates of metrics values, as results of multiple iterations are averaged.

Cross validation measures the performance to detect overfitting and generalization of model on unseen data., values of metrics under study and the variation in the corresponding values of cross-validation folds. Logistic Regression (LR) model accomplished accuracy of 96% and with a (ROC-AUC) score of 0.99. Moreover, 97% precision for real news and 96% for fake news, alongside recall values of 96% for real news and 97% for fake news. The F1-Score, along with the balanced accuracy of 96%, further confirm the model's balanced performance. In addition to it, the Matthews Correlation Coefficient (MCC) and Cohen's Kappa, contains values 0.9379 and 0.9372 respectively, indicates strong correlation between predicted and actual classifications, while high value of specificity also underscore the model's reliability in distinguishing true negatives. Overall, these metrics collectively highlights effectiveness and consistency of classifier in fake news detection.

Cross validation results revealed Logistic Regression model (LR) perform better than Navie Byes (NB) on unseen data set and surpasses Navie bayes (NB) in terms of recall, precision, accuracy. etc., depicts logistic model's performance is noteworthy in identification of fake news.

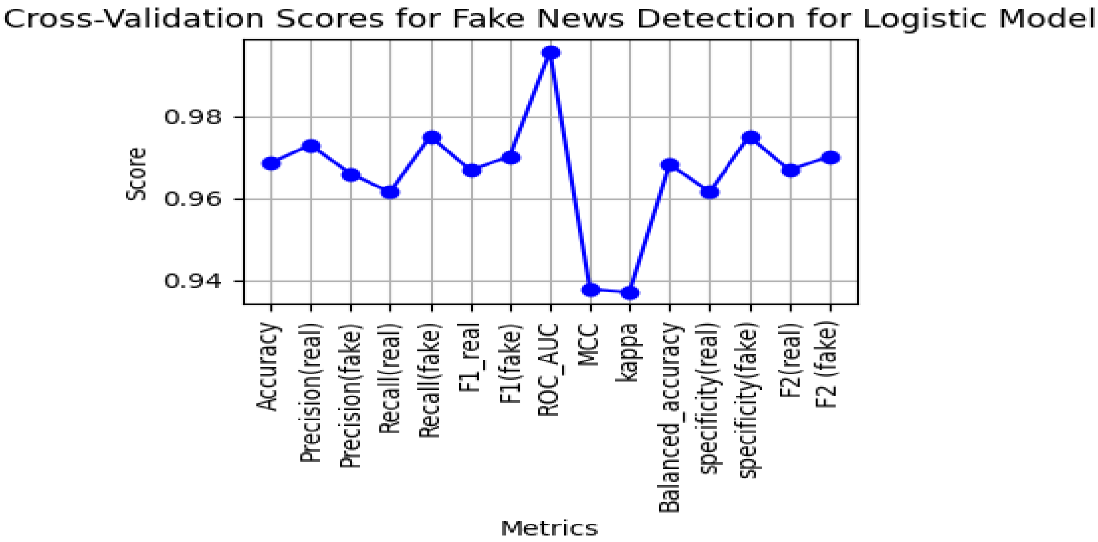

Figure 7. Plot of Cross validation for Logistic Regression (LR) model.

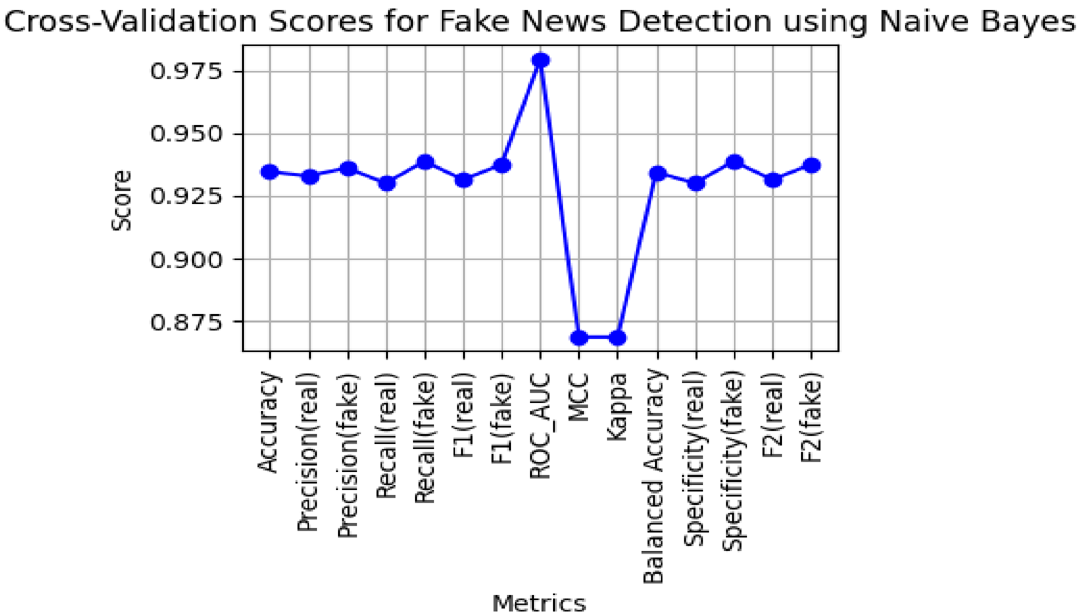

Figure 8. Plot of cross validation of matrices for Navie Bayes (NB)model.

In figures 7 & 8, cross-validation plots show that our fake news detection models perform well across a wide range of metrics, indicating their robustness and reliability. The high scores in precision, recall, F1-Score, and (ROC-AUC) metrics suggest that the model is effective and has ability to distinguish between real and fake news. The low standard deviations depict less variation and highlight the of model showing consistency across different cross-validation folds. This plot also illustrates the uniformity between different evaluated metrics. After comparative study it is concluded that logistic model proved to be better in perspective of slightly higher values of metrics during cross validation.

Conclusion

In this study Comparative analysis between Logistic Regression (LR) model and Navie Bayes (NB) classifier was conducted and our findings indicate that logistic model Performs remarkably well and attained outstanding accuracy and precision of 98% for 8980 given instances of test data in classifying fake and true news articles based on the given thresholds set values. Precision, Specificity, F1-Score, F2-Score and recall, all are closer to 1 supports the good performance of suggested model for detection of fake news with error approximately of 2%. (MCC) and Cohen's Kappa score further highlights the model's strong predictive performance, and empathies on correlation between predicted and actual values of metrics in comparison to Navie Bayes (NB) classifier obtained accuracy of 93% and it trailed behind in terms of values all metrics. Moreover, Balanced accuracy and specificity describes the logistic model's effectiveness for imbalance classes of fake and real news. The F2- Score affirms the capability of models to give priority to recall, which is very important, where predicting fake news (false negatives) is more valuable than wrongly predicting real news (false positives). However, its probabilistic outputs do not effectively distinguish between the classes, as indicated by an (AUC) of 0.002 and a straight-line (ROC) curve, despite these high metrics, the capability of model to differentiate between two classes of fake and real news based on predicted probabilities, was not noticeable, explained by an Area Under the Curve (AUC) and a straight-line Receiver Operating Characteristic (ROC) curve. This is clear indication of logistic model's binary predictions are robust, which can be improved in future study by optimizing the data set including more articles. This extensive review assures Logistic Regression (LR) model proved to be a better and reliable model in identification of fake news relative to Performance reviewed in current analysis it is due to the fact that Navie bayes (NB) supported the independency between features which is not the case of real-world problems. In conclusion, the logistic model's accuracy and dependability have made it appropriate for practical applications in identifying false news, although there is still opportunity for additional development in the model.

References

1. Ahmad, Iftikhar, Muhammad Yousaf, Suhail Yousaf, and Muhammad Ovais Ahmad. 2020. "Fake News Detection Using Machine Learning Ensemble Methods." *Complexity* 2020. https://doi.org/10.1155/2020/8885861.
2. Aldwairi, Monther, and Ali Alwahedi. 2018. "Detecting Fake News in Social Media Networks." In *Procedia Computer Science*, 141:215–22. Elsevier B.V. https://doi.org/10.1016/j.procs.2018.10.171.
3. Allcott, Hunt, and Matthew Gentzkow. 2017. "Social Media and Fake News in the 2016 Election." *Journal of Economic Perspectives*. American Economic Association. https://doi.org/10.1257/jep.31.2.211.
4. Deokate, Suchitra B. 2019. "Fake News Detection Using Support Vector Machine Learning Algorithm." *International Journal for Research in Applied Science and Engineering Technology* 7 (7): 438–44. https://doi.org/10.22214/ijraset.2019.7067.
5. Gupta, Himank, Mohd Saalim Jamal, Sreekanth Madisetty, and Maunendra Sankar Desarkar. n.d. "A Framework for Real-Time Spam Detection in Twitter."
6. Hussain, Shakira, Naveed Sheikh, Misbah Anjum, Arbab Ghulam Raza, and Rabail Rizvi. 2023. "Mathematical Modelling of COVID-19 Pandemic in Pakistan with Optimal Control." *Journal of Asian Scientific Research* 13 (1): 28–44. https://doi.org/10.55493/5003.v13i1.4721.
7. Jwa, Heejung, Dongsuk Oh, Kinam Park, Jang Mook Kang, and Heuiseok Lim. 2019. "ExBAKE: Automatic Fake News Detection Model Based on Bidirectional Encoder Representations from Transformers (BERT)." *Applied Sciences (Switzerland)* 9 (19). https://doi.org/10.3390/app9194062.
8. Saeed, Saeeda, Junaid Baber, Maheen Bakhtyar, Ihsan Ullah, Naveed Sheikh, Imam Dad, and Anwar Ali Sanjrani. 2018. "Empirical Evaluation of SVM for Facial Expression Recognition." *IJACSA) International Journal of Advanced Computer Science and Applications*. Vol. 9. www.ijacsa.thesai.org.
9. Saleem, Muhammad, Naveed Sheikh, Abdul Rehman, Muhammad Rafiq, and Shah Jahan. 2023. "Real-Time Object Identification Through Convolution Neural Network Based on YOLO Algorithm." *Mathematics and Computer Science*, December. https://doi.org/10.11648/j.mcs.20230805.11.
10. Sherazi, Kalsoom, Naveed Sheikh, Misbah Anjum, and Arbab Ghulam Raza. 2023. "Solar Drying Experimental Research and Mathematical Modelling of Wild Mint and Peach Moisture Content." *Journal of Asian Scientific Research* 13 (2): 94–107. https://doi.org/10.55493/5003.v13i2.4814.
11. Shu, Kai, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu. n.d. "Fake News Detection on Social Media: A Data Mining Perspective." http://www.journalism.org/2016/05/26/news-use-across-.
12. Shu, Kai, Suhang Wang, and Huan Liu. 2019. "Beyond News Contents: The Role of Social Context for Fake News Detection." In *WSDM 2019 - Proceedings of the 12th ACM International Conference on Web Search and Data Mining*, 312–20. Association for Computing Machinery, Inc. https://doi.org/10.1145/3289600.3290994.
13. Vedova, Marco L Della, Eugenio Tacchini, Stefano Moretécole, Gabriele Ballarin, Massimo Dipierro, and Luca De Alfaro. n.d. "Automatic Online Fake News Detection Combining Content and Social Signals." https://developers.facebook.com/docs/graph-api.
14. Wu, Tingmin, Shigang Liu, Jun Zhang, and Yang Xiang. 2017. "Twitter Spam Detection Based on Deep Learning." In *ACM International Conference Proceeding Series*. Association for Computing Machinery. https://doi.org/10.1145/3014812.3014815.
15. Yang, Shuo, Kai Shu, Suhang Wang, Renjie Gu, Fan Wu, and Huan Liu. n.d. "Unsupervised Fake News Detection on Social Media: A Generative Approach." https://www.cnbc.com/2016/12/30/read-all-about-it-the-.