Detecting Deceptive Reviews in E-commerce: A Machine Learning Approach

Ayyapasetti Abhinav
Department of Computer Science and Engineering
Amrita Vishwa Vidyapeetham
Chennai, India
abhiayyapasetti@gmail.com

***Abstract*—Every person in the world is using the internet, this has made E-Commerce grow. People after using a new product or after visiting a new hotel, are interested in posting reviews about their experience. After visiting a newly started hotel the owner may ask to post a positive review on them, as they have started recently. Even if the experience is not good the people may post fake reviews based on their different conditions. Another person who will book a hotel based on reviews may get cheated. So it is very important to detect deceptive reviews. This research paper has developed a model that can detect deceptive reviews. Support vector machine and Gradient Boosting Classifier machine learning models have performed very well in terms of accuracy, precision, and recall with 100%.**

***Index Terms*—Natural language processing, Sentimental analysis, Natural language Toolkit, Deceptive, Count vectorizer, Support Vector Machine**

I. Introduction

In the new age, the use of E-commerce has increased. People post reviews after using a new product or visiting a new place. Based on reviews posted new customers decide to buy a new product or visit a new place. People used to rely on recommendations from friends or travel agencies when choosing a place to stay. Now, they turn to online reviews. These reviews are like personal recommendations from other travelers who've stayed in those hotels. Businesses can benefit greatly financially from positive reviews, but negative reviews can have a negative impact. Reading about someone else's experience helps to know if a hotel is as good as it claims to be. If a lot of positive reviews have been posted, it makes the person reading reviews feel more confident about the choice of booking a hotel. But with this trust comes a challenge. Some people write fake reviews to trick people into making the wrong choice. These fake reviews can be like traps, making bad hotels seem good. Identifying fake reviews has become crucial. The big challenge is distinguishing fake reviews from real reviews.

Some companies turn to creating fake favorable reviews to boost their reputation, draw in more clients, and eventually boost sales. People with personal grudges against a company or its owners may create fake negative reviews to avenge them. The current issue is demonstrating to people how to distinguish real reviews from fake ones. Reviews are frequently used by tourists to learn more about the caliber of lodging, services, and overall experiences. Maintaining the integrity of the online review ecosystem now depends on finding and eliminating fraudulent ratings on travel websites. The current study aims to create an intelligent system that can identify fake hotel booking website reviews.

II. Literature Survey

S N Alsubari[1] has proposed a machine-learning approach that uses the gold standard dataset and mainly focuses on feature extraction. The authors used TF-IDF for feature extraction. They used n-gram features to represent text data, here four-grams are used. Before feature extraction, authors used tokenization to divide the text into individual words, and then parts of speech tagging were applied to split words. The author proposed four supervised machine-learning algorithms Support Vector Machine, Adaptive Boosting, Random forest, and Naive Byes achieved accuracies of 0.95, 0.94, 0.93, and 0.88 respectively. The Random Forest has achieved 0.95 accuracy and F1 score metric.

Zulpan Hadi[2] has proposed detecting fake reviews using a Support Vector machine and Random Forest. The dataset used is from kaggle named Tokopedia Product Reviews. The authors focus mainly on data preprocessing using case folding to lower case text, tokenization to split words, stop-word to remove meaningless words, stemming from converting words to their root form, and Part of speech tagging. Support Vector Machine has performed better than the Random Forest by getting an accuracy of 0.98 accuracy.

Ahmed M. Elmogy[3] used the Yelp dataset which consists of restaurant reviews. The authors aimed to identify fake reviews from real ones focusing on data preprocessing. They used data preprocessing techniques such as Tokenization, Stop Words Cleaning, and Lemmatization. During feature extraction, authors used Cosine similarity, n grams and TF-IDF. In feature engineering, authors added features such as the total number of capital letters, the total number of punctuations, and the total number of emojis in every review. The author proposed Random Forest, Support Vector Machine, Naive Bayes, KNN, and Logistic regression machine learning models. Here, the KNN machine learning model with k value 7 performs better with a 0.824 f1 score in bigrams and 0.822 in trigrams.

Dr. Chandaka Babi[4] proposed a deep-learning BERT(Bidirectional Encoder Representations from Transformers) model for feature extraction. The authors used the Twitter reviews dataset. They added features such as semantic polarity, word count, and review text length to the dataset. The dataset is divided into 70:30 train and test split. The proposed Support Vector Machine model got an accuracy of 1 performing better than Naive Bayes which got an accuracy of 0.97.

Another text-based model was developed by A B Hari Krishnan[5] who used Gold Standard Dataset. Data preprocessing techniques tokenization and lemmatization are used. Bag of Words and TF-IDF are used to convert text into numeric values. To avoid the risk of overfitting, K-fold cross-validation (k=5) is used. The dataset is divided into 5 subsets. Max feature in Bag of Words played a big role in affecting accuracy. The author also used deep learning models like RoBERTa, BERT, LSTM, and CNN. RoBERTa deep learning model with entire data augmentation has outperformed other models with an accuracy of 1 in the training set and 0.998 in the test set.

Abrar Qadir Mir[6] has proposed a Support Vector Machine model that detects deceptive reviews. The author used a dataset which is formed by merging four datasets. This dataset is called the gold standard mock review dataset. Here, word embeddings were extracted from review text using Bidirectional Encoder Representation from Transformers(BERT). This dataset is evaluated by machine learning models Support Vector Machine, Bagging Classifier, Adaptive Boosting, Naive Bayes, KNN. SVM outperforms them with an 0.88 F1 score.

III. DATASET

This paper used a dataset named Gold Standard Dataset which was created by Ott[7]. This dataset has 1600 instances of reviews and 5 attributes which are collected from sources TripAdvisor, MTurk, and Web. The dataset consists of the following features:

- **Deceptive**
  Feature contains the classification of review text whether it is truthful or deceptive.
- **Hotel**
  Provides hotel name, this feature contains popular 20 hotel names in Chicago.
- **Polarity**
  The polarity feature describes the sentimental analysis of the review text: Positive and negative.
- **Source**
  Provide the name of the source where the review is taken from.

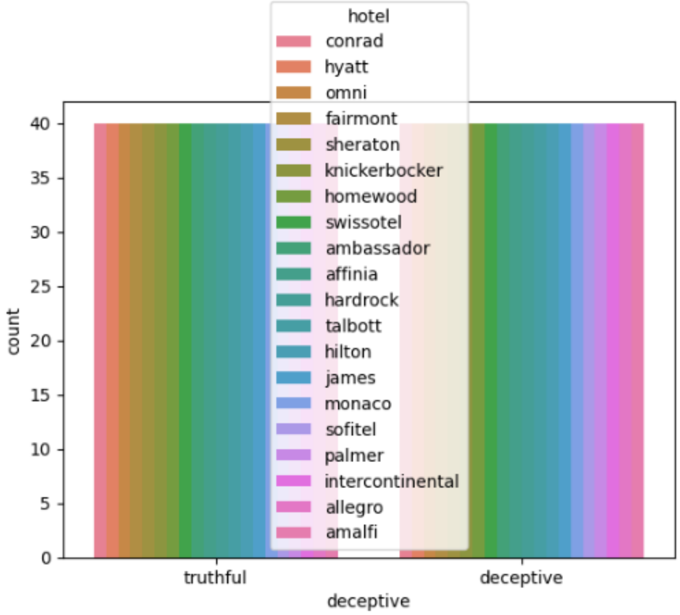

Fig. 1. Name of the Hotel

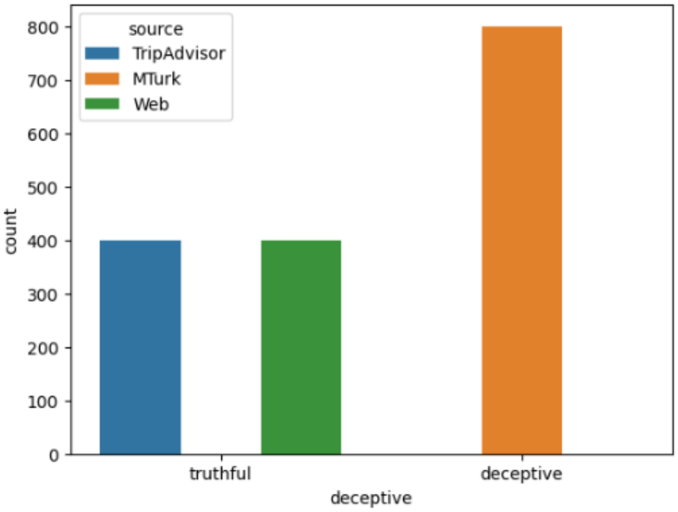

Fig. 2. Reviews collected from Sources

- **Text**
  The user-written review text data.

The dataset was broken down into 400 deceptive positive reviews,400 truthful positive reviews, 400 deceptive negative reviews, and 400 truthful negative reviews. This dataset consists of a balanced collection of deceptive truthful and reviews, both positive and negative. As the dataset is totally text data using Natural Language Processing the raw textual data can be transformed into structured data. Effective textual data processing and analysis are made possible by Natural Language Processing(NLP) approaches.

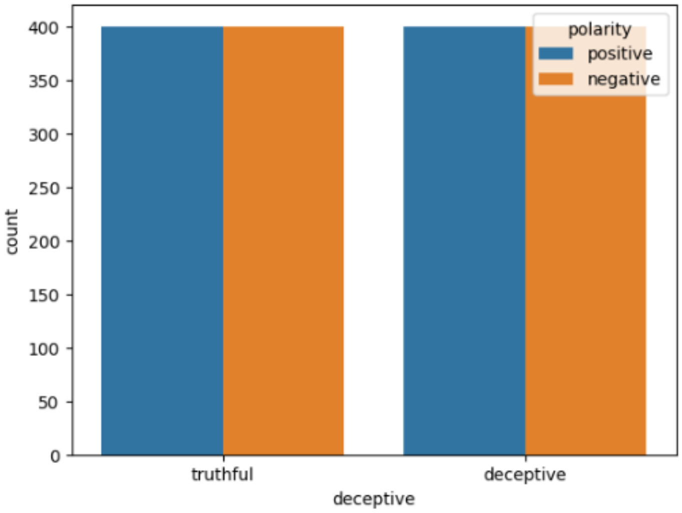

Fig. 3. Balanced collection of Dataset

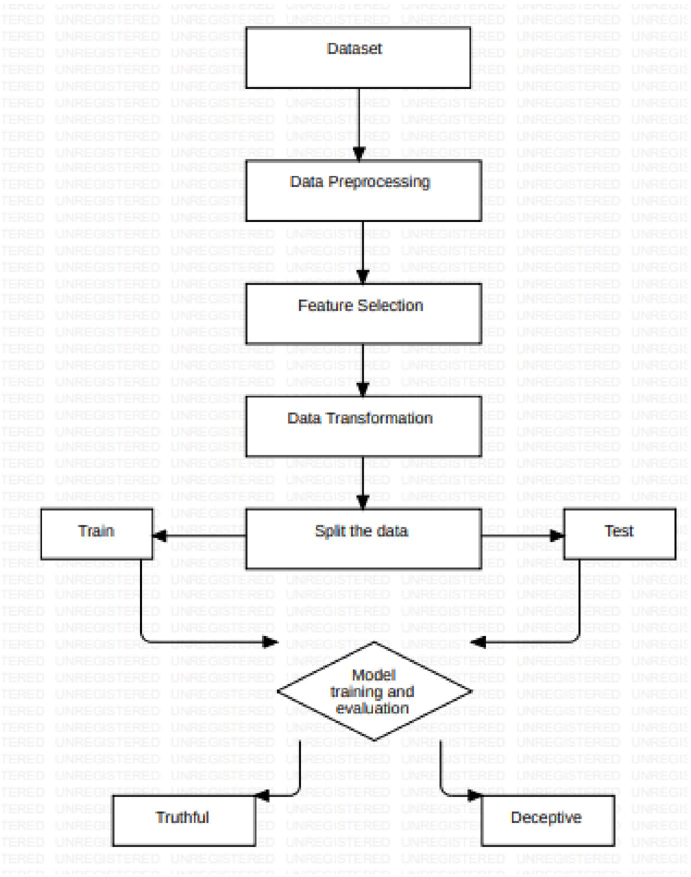

Fig. 4. Methodology Flowchart

IV. Methodology

A. Data Preprocessing

The dataset is completely text so it is very important to handle the text data. The dataset must be applied to certain cleanings. As the text is a human language, Natural language processing(NLP) techniques can be used to make the text understood by computers. NLP can be used to extract specific meanings from the text. Text data cannot be given directly to evaluate with a machine learning model. Using the NLP techniques the text can be converted to numeric values. Based on the review the sentimental score can also be checked whether the review is positive or negative, this can be extracted using NLP techniques. The model needs to perform well in training data as well as in training data. Overfitting occurs if the model produces high accuracy in train data but less accuracy in test data. If this happens then while evaluating real-world data, the model will definitely not perform well. When both train and test data do not do well underfitting occurs.. Using the Natural Language Toolkit library in Python the NLP techniques can be implemented on text data.

1) Data Cleaning: Firstly, the dataset has been checked whether it contains any null values. There are no null values present in the dataset. A new feature complete_text has been added to the dataset which contains the combination of feature source and feature text. Feature complete_text has been applied for data cleaning. Applied case folding converting all the text to lowercase because the NLTK treats the same words with uppercase and lowercase differently. Using the Regular Expressions library certain changes in the complete_text feature have been made. Removed text that is present in square brackets. Removed special characters. Removed links present in complete_text feature. Removed HTML tags. Removed punctuations in the text. Removed the new line characters. Removed words that contain numbers. Data cleaning has been done by applying all these to the text_cleaning feature in the dataset.

2) Tokenization: Using tokenization splitting the cleaned complete_text into individual words has been done. Tokenization has been done to give these tokenized words to check for common stop words.

3) Stop Words: In NLTK common stopwords for the language, English will be available. So when the tokenized words are given as input, whenever a stop word is detected it will be removed from the complete_text feature. Stop words are removed because they do not contain much value as compared to other words. They have no meaning. At last, the tokenized words are combined to form a string.

4) Stemming: Converts words present in complete_text feature to their root form using PorterStemmer library present in NLTK library. By doing this the complete_text can be consistent. For example, stay and stayed are the same words but treated differently by after stemming they can be treated the same. This helps in decreasing the complexity of vocabulary. So sentimental analysis can be easily done.

B. Feature Selection

From the dataset deceptive feature is selected as the target variable Y. complete_text feature is taken as variable X. Remember these column values are still in text data. These text values cannot be given directly to a model.

C. Data Transformation

Using Count vectorizer the X is converted into numeric values. Here in the vectorizer both unigram and bigram were used for ngram range. Y contains categorical target features so label encoding is used for converting into numeric values. Now, the X and Y feature numeric values.

D. Split the data

By using the sci-kit learn library splitting the X and Y into 0.8 X_train, Y_train, and 0.2 X_test and andY_test. Splitted the data into train and test.

Data preprocessing, feature selection, Data transformation, and Data split have been completed so now it is ready to be evaluated by the machine learning model.

E. Model training

1) Support Vector Machine: Support Vector Machine has a hyperplane that classifies different classes. This dataset is a binary classification. Using SVM can avoid overfitting because it has margin maximization between data points to the hyperplane. Support Vector Machine has many kernels. In this paper Linear kernel, RBF kernel, and Sigmoid kernel were checked. Linear kernel has achieved 100% accuracy, precision, and recall with no false trues and false positives.

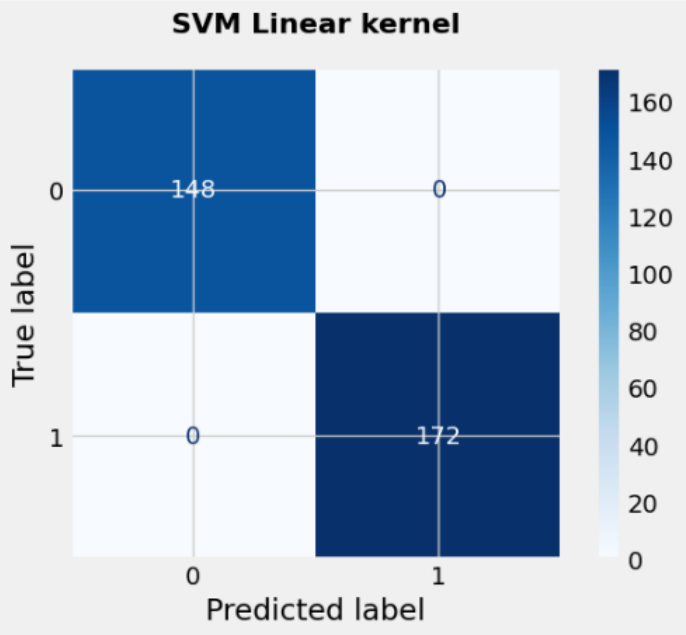

Fig. 5. Confusion matrix for SVM linear kernel

2) Naive Bayes: Naive Bayes is a probabilistic model. Naive Bayes Classification is done based on likelihood, Prior, and posterior probabilities. From the probability, the Naive Bayes model classifies it into a specific category. In this paper multinomial NB is used which is imported from the sci-kit learn naive Bayes library. Naive Bayes achieved 0.92 accuracy, 0.93 precision, and 0.93 recall.

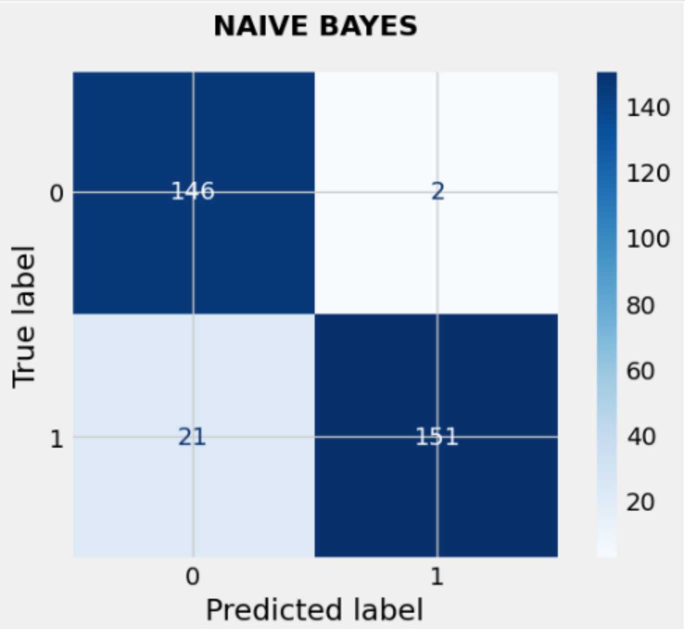

Fig. 6. Confusion matrix for NB

3) Random Forest: Random Forest mainly focuses on reducing overfitting. It is a combination of many decision trees. Decision trees are built by checking features which is getting high information gain. Random forest is imported from the ensemble sci-kit library. Like this many decision trees are combined for a random forest model this is why it is called an ensemble method. Random Forest achieved a 0.99 accuracy, 0.99 precision, and 0.99 recall.

4) Extra Trees Classifier: Extra Trees Classifier is an ensemble method. Extra trees are the combination of many decision trees. As we have seen in Decision trees which use best split to select a feature, here random features will be selected because the Extra Trees Classifier has a randomness feature. This is imported from the ensemble library. Randomness can be adjusted using random_state. Extra Trees Classifier achieved 99.0625% accuracy and precision, recall values of 0.99, 0.99 respectively.

5) Gradient Boosting classifier: Gradient Boosting classifier builds a strong model by combining all the weak learners. In this paper, the weak learner used is decision trees. The

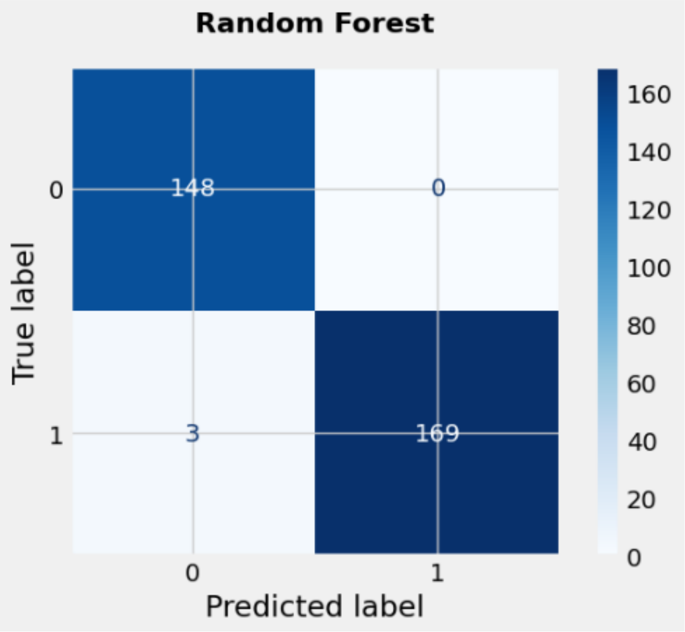

Fig. 7. Confusion matrix for RF

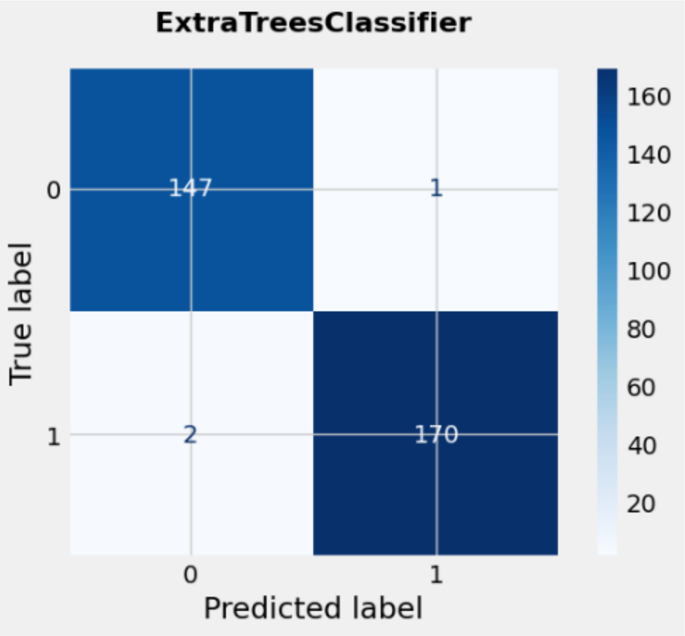

Fig. 8. Confusion matrix for Extra Trees Classifier

number of weak learners can be set using n_estimators. In this paper, it is set to 100 n_estimators. This model is imported from the Sci-kit Learn ensemble library. This model achieved 1 accuracy, precision, and recall.
All these machine learning models false true and false negatives were represented in the confusion matrix.

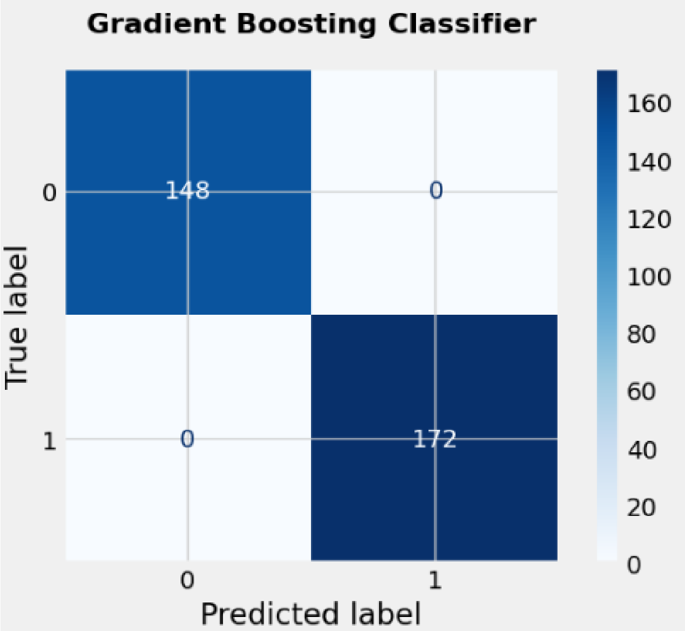

Fig. 9. Confusion matrix for Gradient boosting Classifier

TABLE I
EVALUATED MODEL RESULTS

Model	**Performance Metrics**			
	Train Accuracy	***Test Accuracy***	***Precision***	***Recall***
Support Vector Machine Linear Kernel	100	100	100	100
Support Vector Machine RBF Kernel	100	99.375	99	99
Support Vector Machine Sigmoid Kernel	99.21	99.68	100	100
Naive Bayes	100	92.81	93	93
Random Forest	100	99.06	99	99
Extra Trees Classifier	100	98.75	99	99
Gradient Boosting Classifier	100	100	100	100

V. CONCLUSION

This research paper talks about how important it is to detect deceptive reviews. Firstly, Data preprocessing has been done because data preprocessing is the most important not only to increase accuracy but also to decrease model detecting false trues or false positives. Here several data preprocessing methods such as data cleaning, tokenization, Stop word removal, and stemming were implemented. Count vectorizer was used to convert data into numeric values which formed a sparse matrix. For categorical values, a label encoder is used. Support Vector Machine(Linear kernel) and Gradient Bosting Classifier have performed very well in terms of test, train accuracies, and also in terms of recall and precision. Support Vector Machine(Sigmoid kernel) has performed very well by detecting no false trues and no false positives. In paper[1] Random Forest achieved an accuracy of 95%. In this research paper increased the Random Forest accuracy to 99.05%. Ensuring that no overfitting or underfitting has occurred. Precision and Recall also got 99%. It is necessary to have a good Recall and Precision not only accuracy.

REFERENCES

[1] Saleh Nagi Alsubari, Sachin N. Deshmukh, Ahmed Abdullah Alqarni, Nizar Alsharif, Theyazn H. H. Aldhyani, Fawaz Waselallah Alsaade and Osamah I. Khalaf6, “Data Analytics for the Identification of Fake Reviews Using Supervised Learning”

[2] Zulpan Hadi1, Ema Utami, Dhani Ariatmanto, ”Detect Fake Reviews Using Random Forest and Support Vector Machine”

[3] Ahmed M. Elmogy , Usman Tariq, Ammar Mohammed , Atef Ibrahim, ”Fake Reviews Detection using Supervised Machine Learning ”

[4] Dr. Chandaka Babi , M. Sai Roshini , P. Manoj , K. Satish Kumar, ”Fake Online Reviews Detection and Analysis Using Bert Model”

[5] Anusuya Baby Hari Krishnan, ”Unmasking Falsehoods in Reviews: An Exploration of NLP Techniques”

[6] Abrar Qadir Mir, Furqan Yaqub Khan, Mohammad Ahsan Chishti, ”ONLINE FAKE REVIEW DETECTION USING SUPERVISED MACHINE LEARNING AND BERT MODEL”

[7] M. Ott, Y. Choi, C. Cardie and J. T. Hancock, “Finding deceptive opinion spam by any stretch of the imagination,”