FOREX Publication
Open Access | Rapid and quality publishing

International Journal of
Electrical and Electronics Research (IJEER)
Research Article | Volume 12, Special Issue -BDF | Pages 28-35 | e-ISSN: 2347-470X

# Taming Misinformation: Fake Review Detection on Social Media platform using Hybrid Ensemble Technique

**Shraddha Kalbhor[1], Dinesh Goyal[2] and Kriti Sankhla[3]**
[1]*Computer Science & Engineering, Poorinma University, Jaipur shraddha.kalbhor000@gmail.com*
[2]*Professor, Computer Science & Engineering, Poorinma University, Jaipur dinesh8dg@gmail.com*
[3]*Associate Professor, Computer Science & Engineering, Poorinma University, Jaipur kriti.sankhla@poornima.edu.in*

***Correspondence:** shraddha.kalbhor000@gmail.com

**ABSTRACT-** In today's digital world, we witness exponential growth in the generation of textual content on a daily basis. However, the widespread dissemination of information through social media, online forums, and news websites has given rise to the proliferation of fake views, opinions, and reviews, posing a significant challenge in the battle against misinformation and manipulation. Machine Learning has become increasingly integral to real-world online activities, particularly in the area of Artificial Intelligence. Traditional methods often struggle to keep pace with the relentless creation of internet data. Consequently, short text processing has emerged as a new domain for the application of Machine Learning. This is where sentiment analysis come to the forefront, offering potent tools for discerning the authenticity of online content. Detecting and combating these fabricated sentiments are crucial for preserving the integrity of information and ensuring informed decision-making. This work focus on the previously unexplored area of user comments on review data. By leveraging N-gram technique and hybrid ensemble classification approaches, the research addresses critical issues in fake reviews classification, and sentiment analysis. The aim of this work is to detect fake reviews with limited text using NLP feature extraction, and hybrid ensemble classification algorithms, ultimately contributing to the enhancement of information integrity and decision-making in the digital age.

**Keywords:** Social media, machine learning, n-gram technique, NLP feature extraction, fake reviews.

**ARTICLE INFORMATION**

**Author(s)**: Shraddha Kalbhor, Dinesh Goyal and Kriti Sankhla;

**Received**: 07/10/2023; **Accepted**: 20/12/2023; **Published**: 28/03/2024;
**e-ISSN**: 2347-470X;
**Paper Id**: IJEER-BDF05;
**Citation**: 10.37391/ijeer.12bdf05
**Webpage-link**:
https://ijeer.forexjournal.co.in/archive/volume-12/ijeer-12bdf05.html

This article belongs to the Special Issue on **Innovations and Trends in Computer, Electrical, and Electronics Engineering: Bridging the Digital Frontier**

**Publisher's Note:** FOREX Publication stays neutral with regard to jurisdictional claims in Published maps and institutional affiliations.

## 1. INTRODUCTION

Today, users heavily rely on online platforms for product purchases, offering convenience, diverse options, competitive pricing, and streamlined delivery processes. The integral aspect of online shopping is further emphasized by consumers' active engagement through reviews, sharing valuable feedback after their purchases. This is because many e-commerce sites collect and store short texts like product reviews, share information, and send emails to our webserver. Because there are so many different kinds of short messages, the need for good leadership has also grown. With traditional ways of processing short texts, it's hard to keep up with the high rate of data creation on the internet, so short text processing is a natural way for Machine Learning to be used.

Social media, online communities, and news websites make it easy for a lot of people to share information, which has led to a lot of fake views, opinions, and reviews. Finding this fake review is a very important part of the fight against fake news and influence. This is where sentiment analysis [1] and aspect-based sentiment analysis come in. These are strong ways to figure out if online content is real or not.

So, fighting against fake views and twisted feelings in online debate is important to keep information honest and make sure people can make good decisions; when combined with aspect-based sentiment analysis. Sentiment analysis gives us the tools to look closely at the emotions expressed in digital material. This helps us find and stop the spread of false views. In this introduction, we start a work to find out how these strong NLP methods can be used in the ongoing fight against fake views and to keep online conversation honest. In this current online and smart learning process, the most widely and consistently used video repository is YouTube. This platform of YouTube provides or allows the users to express their views, emotions etc. about the respective videos in the form of comments and ratings. These user comments are still unexplored in the context of ranking in retrieval. The only criterion to decide whether a video is relevant and also to decide its quality is based on the like or dislike option of the video, which in turn also decides its ranking. But sometimes there are cases where a highly rated video shows no relevancy to the topic which one comes to know only after watching it.

Hence our work helps researchers in the fields of fake reviews classification and sentiment analysis and the spread of learning to get around a number of problems. The work uses ML, DL, and hybrid classification approaches and perform Multi-Class Classification on text dataset.

### 1.1 Proposed Contribution

- The goal is to create a system that can properly detect fake social media reviews and assist consumers in avoiding being deceived.
- Leverage the power of N-gram methods to identify fake reviews. The work intends to improve the accuracy and efficacy of fake review categorization by using N-gram-based feature extraction approaches, particularly in settings with limited text.
- Finally, hybrid ensemble classification techniques will be utilised to determine if the reviews are fake or real.

## 2. LITERATURE REVIEW

Numerous studies highlight challenges in detecting fake reviews, emphasizing NLP techniques, such as text preprocessing and sentiment analysis. Akhtar et al. [2] employed machine learning for fake review detection, revealing a gap in single-objective optimization. Manek et al. [5] achieved impressive sentiment analysis results but identified areas for further exploration. Basiri et al. [12] focused on bias detection, while Li et al. [21] excelled in text classification. Kermani et al. [20] conducted mood analysis, suggesting potential for advanced methodologies. Ceron Andrea et al. [7] demonstrated supervised sentiment analysis effectiveness in electoral monitoring. Z. Kastrati et al. [9] explored weakly supervised learning for aspect-based sentiment analysis, highlighting efficiency challenges. Lv et al. [10] introduced CAMN for mood analysis, Deng et al. [18] stressed domain-specific sentiment lexicons, collectively showcasing the evolving landscape of sentiment analysis methodologies. Pranckevicius Tomas et al. [6] evaluated traditional ML models for sentiment analysis, revealing unexplored potential in advanced and deep learning techniques. Al-Shammari et al. [14] focused on fake news detection, Siering et al. [19] emphasized expanding sentiment analysis in tourism. Huang et al. [17] explored multi-label learning, Fan et al. [3] introduced MGAN for sentiment analysis, while Yang et al. [8] and Yan et al. [11] presented SLCABG and CNN-BiGRUAT models, respectively. Ishaq et al. [15] leveraged CNN, semantic feature mining, and Word2Vec, and Ghani et al. [16] addressed memory-intensive lazy learning methods, collectively enriching sentiment analysis and its applications.

## 3. PROPOSED WORK

### 3.1 Proposed System

Detecting fake reviews in the age of abundant online content is a critical challenge, and several key steps are integral to this process. Firstly, data collection from kaggle is use as input for analysis. Pre-processing steps like text cleaning, tokenization, and stop-word removal help ensure data quality. Feature extraction techniques, including NLP-based approaches, are then applied to capture meaningful information from the text. One crucial innovation in this domain is the proposed hybrid ensemble classifier, which combines multiple classification algorithms. This ensemble approach leverages the strengths of different models, enhancing the accuracy and robustness of fake review detection. By integrating techniques such as N-grams and sentiment analysis, this hybrid classifier offers a comprehensive solution to identify fabricated sentiments in reviews. *Figure 1* shows the proposed architecture of fake review detection system and description of important steps is given below;

#### 3.1.1 Input Dataset

The dataset consists of reviews from various domains, including restaurants, laptops, and hotels. Each review is annotated with specific aspects or targets that are being evaluated, along with the corresponding sentiment expressed towards each aspect. The sentiment labels can be positive, negative, or neutral, indicating the overall sentiment associated with the given aspect.

#### 3.1.2 Dataset Pre-Processing

In the data processing pipeline for our review dataset, we initiated by calculating the word count for each review, providing us with valuable insights into the length and verbosity of the text data. Next, we identified unique keywords within the corpus, allowing us to understand the diversity of vocabulary used by reviewers. To ensure data integrity, we performed Nan (missing value) removal, effectively eliminating any incomplete or irrelevant entries that might skew our analysis. Lastly, we conducted word frequency analysis to determine the most commonly occurring terms across the dataset, facilitating the identification of important keywords and trends within the reviews. These comprehensive data processing steps not only prepared our dataset for subsequent analysis but also provided essential statistics and insights crucial for further exploration and modelling.

#### 3.1.3 Data (Review) Normalization

Text normalization techniques such as lowercase conversion, punctuation removal, stemming, lemmatization, stopwords removal, acronym substitution, and contraction substitution are commonly used on review datasets for several reasons:

- *Lowercasing*: Transforming any text into lowercase standardises the content and simplifies word matching. It ensures that words with different capitalization are treated as the same word. For example, "Good" and "good" will be treated as the same word after lowercase conversion.
- *Punctuation Removal*: Eliminating noise and superfluous symbols from the text can be achieved by removing punctuation indications, such as question marks, exclamation marks, periods, and commas.
- *Stemming*: The process of stemming involves taking words' prefixes and suffixes off in order to reduce them to their base or root form. Stemming, for example, would reduce

Research Article | Volume 12, Special Issue -BDF | Pages 28-35 | e-ISSN: 2347-470X

"studying", "studies" and "studied" to the basic form "study".

- *Lemmatization*: It is similar to stemming, but instead of just turning a word into its root form, it creates a new word called a lemma. It uses the word's context and grammatical study to change it to its base form. Lemmatization makes sure that the new word makes sense and can be looked up in a dictionary.

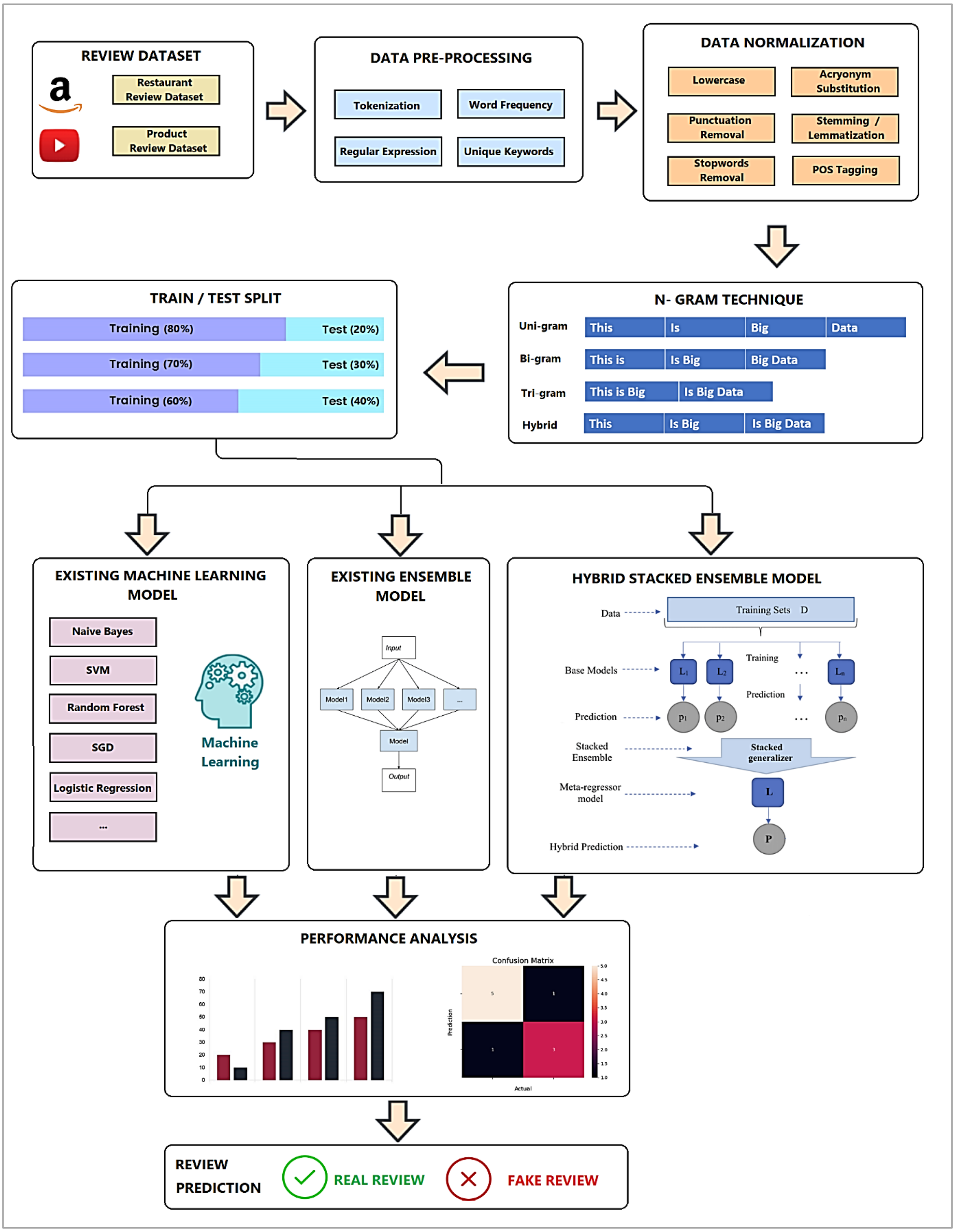


**Figure 1:** Proposed System Architecture

FOREX Publication
Open Access | Rapid and quality publishing

International Journal of
Electrical and Electronics Research (IJEER)
Research Article | Volume 12, Special Issue -BDF | Pages 28-35 | e-ISSN: 2347-470X

- Stopwords removal: stopwords are familiar words like "a," "the," "is," "and," etc., that are used a lot but don't add much to the sense of the sentence as a whole.
- Substituting Acronyms: Acronyms are often used in online reviews. Using their full names instead of the acronyms helps keep the sense and readability of the text. For example, changing "LOL" to "laugh out loud" makes it easier to understand what the person is trying to say.
- Contraction Substitution: Extending contractions to their full forms is what contraction substitution is all about. For example, use "do not" instead of "don't" or "can't" instead of "cannot." This makes sure that words are always treated the same way and helps keep the right understanding of what they mean.

By applying these text normalization techniques, the review dataset becomes cleaner, more consistent, and easier to process. The transformed text is better suited for analysis, natural language processing tasks, and machine learning algorithms that rely on textual data.

### 3.1.4 N-Gram Technique for Text Processing (Review)

N-grams play a crucial role in capturing local context, improving language modeling, extracting features, handling out-of-vocabulary words, mitigating sparsity, disambiguating word meaning, and supporting various NLP tasks and applications. N-gram technique is a fundamental method in text processing that involves breaking down a piece of text into contiguous sequences of N words or characters. These sequences, called "n-grams," can be unigrams (single words), bigrams (two-word sequences), trigrams (three-word sequences), or higher-order n-grams, depending on the chosen value of N. N-grams capture the local context and relationships between words or characters within a text, making them useful for various NLP tasks such as language modelling, text generation, and sentiment analysis.
Here we have use Uni-gram, bi-gram and tri-gram along with fusion of unigram and bigram features for achieving better performance.

| positive | | | negative | | | neutral | | |
|---|---|---|---|---|---|---|---|---|
| | Words | Frequency | | Words | Frequency | | Words | Frequency |
| **634** | food | 621 | **398** | food | 206 | **365** | food | 131 |
| **1505** | service | 359 | **916** | service | 137 | **386** | get | 75 |
| **1237** | place | 206 | **696** | order | 93 | **267** | dinner | 69 |
| **1280** | price | 204 | **749** | place | 87 | **620** | order | 69 |
| **1831** | well | 182 | **424** | get | 70 | **396** | go | 50 |
| **694** | go | 112 | **785** | price | 69 | **670** | place | 44 |
| **1234** | pizza | 111 | **849** | restaurant | 56 | **73** | bar | 44 |
| **1840** | wine | 106 | **1052** | time | 48 | **1015** | wine | 41 |
| **1009** | menu | 104 | **623** | menu | 44 | **284** | drink | 40 |
| **963** | make | 101 | **771** | portion | 41 | **754** | restaurant | 40 |

**Figure 2.** Unigram Features

### 3.1.5 Train – Test Split

The choice of train-test split ratios, whether it be 80% - 20%, 70% - 30% plays a pivotal role in the machine learning workflow and can significantly impact model performance. A larger training dataset, such as the 80% - 20% split, often provides more data for the model to learn from, potentially leading to better overall performance and generalization. However, it may also leave a smaller portion for testing, making it harder to detect overfitting or assess model robustness. On the other hand, a 70% - 30% split offers a more substantial portion of data for testing, making it easier to evaluate model performance and detect issues like overfitting. It does, however, lessen the quantity of data that is available for training, which may restrict the model's capacity to pick up intricate patterns. The choice of split ratio should be made based on the specific dataset, problem, and available data, striking a balance between training data volume and testing data reliability to ensure robust model development and evaluation. Here we have use all 3 train-test split ratio to see the behavior of proposed model.

## 3.2 Proposed Algorithms

A stacked ensemble hybrid classification model is a type of machine learning that uses the strengths and differences of several different classifiers to improve the general accuracy of predictions. This ensemble model has many layers. Each layer is made up of different classes that are learned separately and then put together in a structured way.

Detailed explanation of the stacked ensemble hybrid classification model:

*Layer 1 - Base Classifiers*: The first layer of the model consists of multiple base classifiers. These classifiers can be of different types, such as DT, SVM, RF, or NN. Each base classifier is trained on the input data using various features and parameters.

*Layer 2 - Meta-classifier*: The predictions from the base classifiers in Layer 1 are then combined as input to a meta-classifier. The meta-classifier is typically a more advanced and powerful model, such as gradient boosting, XGBoost, or a neural network. It learns from the predictions of the base classifiers to make the final classification decision.

*Feature Engineering*: Along with the base classifiers, feature engineering techniques can be applied to the input data. These techniques involve transforming or creating new features that might enhance the predictive power of the models. Common feature engineering techniques include normalization, scaling, one-hot encoding, dimensionality reduction, or creating interaction terms.

*Training and Validation*: The stacked ensemble hybrid model is trained using labeled data. The training data is divided into multiple subsets, and the base classifiers in Layer 1 are trained on these subsets. The predictions from the base classifiers are then used to train the meta-classifier in Layer 2. Cross-validation techniques can be employed to assess the model's performance and prevent overfitting.

*Prediction and Aggregation*: During the prediction phase, unseen data is passed through the base classifiers in Layer 1, and their predictions are obtained. These predictions are then combined using the meta-classifier in Layer 2 to produce the final classification output.

### 3.2.1 Hybrid Ensemble Algorithm

A hybrid ensemble stacking technique for classification tasks. Stacking stands as an ensemble learning technique that amalgamates predictions from numerous base models to enhance the collective predictive performance. It adheres to the subsequent steps:

1. *Model Selection*: Different models, including RandomForestClassifier (entropy / gini), XGBClassifier, DecisionTreeClassifier.
   These models offer diverse learning algorithms and characteristics.
2. *Encoding the Target Variable*: The Label Encoder is used to encode the categorical target variable (y) into numeric labels.
3. *Stacking*: The list of selected models, along with the training features (X_fit_transform_1) and the encoded target variable (label_y), are passed as inputs.
   The stacking function employs cross-validation, dividing the data into multiple folds (5 in this case), and trains each base model on different subsets.
   Out-of-fold predictions and bagging are used to create stacked features from the base models' predictions.
   Stacked features are obtained by combining the predictions of multiple base models.
4. *Meta-Model Training*: A meta-model is trained using the stacked features as inputs and the encoded target variable (label_y). The meta-model employed in this context is a Random Forest (RF) Classifier. This meta-model is tasked with assimilating the predictions from the base models, effectively capturing the combined knowledge of the ensemble.
5. *Making Predictions*: The trained meta-model is used to make predictions on new, unseen data (S_test).
   The meta-model takes the stacked features as input and generates the final predictions for the target variable.
6. *Evaluating the Predictions*: The accuracy score is computed through the comparison of predicted labels (y_pred) with the original labels (label_y).

Through the process of stacking, which involves merging predictions from various models, the hybrid ensemble model strives to enhance the accuracy and resilience of the final predictions in contrast to relying solely on individual models.

### 3.2.2 Machine Learning Technique

Evaluating various machine learning algorithms' performance in sentiment classification is the aim of [4] [13] and identify those yielding the highest performance metrics. By comparing their performance, the most suitable candidates are chosen for a hybrid stacked ensemble classifier. Each algorithm undergoes training and evaluation based on metrics like accuracy, and F1 score. The top-performing algorithms, considering various metrics, are integrated into the ensemble. This selection aims to leverage their strengths for improved mood analysis, creating a robust and dependable system. In this study, we have employed LR, KNN, DT, Linear SVM, RF, and SGD based ML methods.

## 4. RESULT ANALYSIS

### 4.1 Dataset Description

Aspect-Based Sentiment Analysis (ABSA), which is also known as SemEval-2014 Task 4, is a dataset that is used to test systems that do aspect-based sentiment analysis. The information is made up of reviews from many different areas, such as hotels, computers, and restaurants.

In each review, there is a note about the exact parts or goals that are being reviewed, as well as the feelings about each one. The aspect's overall emotion is positive, negative, or neutral. The SemEval-2014 Task 4 dataset for Dataset has Review Text, The Aspect Category/Target, Label.

### 4.2 Performance Parameters

True Positives (TP) - Representing values correctly predicted as positive. True Negatives (TN) - Denoting values accurately predicted as negative. False Positives (FP) – Occurring when the actual class is no, but the predicted class is yes. False Negatives (FN) – Arising when the actual class is yes, but the predicted class is no.

$$Precision(Pre) = \frac{TP}{TP + FP} \quad (1)$$

$$Accuracy(Acc) = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

$$Recall(Rec) = \frac{TP}{TP + FN} \quad (3)$$

$$F1 - Score = 2 * \frac{Pre * Rec}{Pre + Rec} \quad (4)$$

### 4.3 Result and Discussion

In our comprehensive analysis of Review Dataset for various ML algorithms, including LR, KNN Classifier, DT, Linear SVM, RF, SGD, and the Ensemble Hybrid Classifier, we assessed their performance using multiple feature sets, including BOW (All Features) - Unigram, BOW (25% Selected Features) - Unigram, BOW (All Features) - Bigram, BOW (25% Selected Features) - Bigram, BOW (All Features) - Trigram, Mixture Model (Selected Features) - Trigram, and Mixture Model (Unigram + Bigram). The results revealed interesting patterns in algorithm suitability for different feature sets. *Table 3* shows the accuracy comparison of all algorithms with various selected features. From *table 1* we can see that proposed Ensemble Hybrid Classifier (EHC) algorithms performs better compare to other machine learning algorithms, achieving accuracy of 87.06 % for Mixture Model (Unigram + Bigram).

FOREX Publication
Open Access | Rapid and quality publishing

International Journal of
Electrical and Electronics Research (IJEER)
Research Article | Volume 12, Special Issue -BDF | Pages 28-35 | e-ISSN: 2347-470X

**Table 1. Performance comparison of Restaurant and Laptop review dataset for various train test split ratio**

| | BOW (All Features) Unigram | BOW (25 % Selected Features) - Unigram | BOW (All Features)- Bigram | BOW (25 % Selected Features) - Bigram | BOW (All Features) - Trigram | Mixture Model (Selected Features) | Mixture Model (Unigram + Bigram) |
|---|---|---|---|---|---|---|---|
| ***RESTAURANTS REVIEW DATASET*** *(80 % - 20 %) Train Test Split* | | | | | | | |
| LR | 0.666667 | 0.631250 | 0.587500 | 0.475000 | 0.583333 | 0.637500 | 0.700000 |
| KNN | 0.706250 | 0.591667 | 0.708333 | 0.643750 | 0.708333 | 0.658333 | 0.706250 |
| DT | 0.714583 | 0.600000 | 0.718750 | 0.725000 | 0.718750 | 0.656250 | 0.718750 |
| Linear SVM | 0.710417 | 0.575000 | 0.718750 | 0.462500 | 0.700000 | 0.566667 | 0.718750 |
| RF | 0.504167 | 0.500000 | 0.483333 | 0.556250 | 0.516667 | 0.483333 | 0.504167 |
| SGD | 0.704167 | 0.627083 | 0.729167 | 0.710417 | 0.714583 | 0.683333 | 0.708333 |
| EHC | 0.844574 | 0.794778 | 0.787458 | 0.837744 | 0.811124 | 0.804125 | 0.864753 |
| *(70 % - 30 %) Train Test Split* | | | | | | | |
| LR | 0.575491 | 0.518889 | 0.490249 | 0.369261 | 0.385443 | 0.229519 | 0.615708 |
| KNN | 0.583006 | 0.583344 | 0.581584 | 0.582194 | 0.581584 | 0.581584 | 0.581584 |
| DT | 0.626204 | 0.611308 | 0.650374 | 0.608327 | 0.653014 | 0.627014 | 0.633243 |
| Linear SVM | 0.637442 | 0.617603 | 0.641233 | 0.385713 | 0.638931 | 0.235341 | 0.648612 |
| RF | 0.509480 | 0.509618 | 0.256709 | 0.315997 | 0.222945 | 0.275619 | 0.515298 |
| SGD | 0.653758 | 0.633987 | 0.659108 | 0.638929 | 0.658025 | 0.632092 | 0.665607 |
| EHC | 0.760528 | 0.783344 | 0.805079 | 0.807448 | 0.812254 | 0.842451 | 0.870622 |
| ***LAPTOP REVIEW DATASET*** *(80 % - 20%) Train Test Split* | | | | | | | |
| LR | 0.620187 | 0.560564 | 0.520783 | 0.427142 | 0.395845 | 0.613576 | 0.661241 |
| KNN | 0.467264 | 0.486005 | 0.425617 | 0.442665 | 0.427311 | 0.500852 | 0.438766 |
| DT | 0.627907 | 0.615352 | 0.594660 | 0.559632 | 0.575407 | 0.622051 | 0.632910 |
| Linear SVM | 0.671506 | 0.648268 | 0.602635 | 0.424175 | 0.582362 | 0.652932 | 0.675406 |
| RF | 0.538334 | 0.518321 | 0.354210 | 0.413124 | 0.265393 | 0.541644 | 0.563783 |
| SGD | 0.659373 | 0.635711 | 0.612381 | 0.572775 | 0.575915 | 0.640713 | 0.672859 |
| EHC | 0.836547 | 0.757812 | 0.745845 | 0.791424 | 0.821454 | 0.831054 | 0.870611 |
| *(70 % - 30%) Train Test Split* | | | | | | | |
| LR | 0.613156 | 0.544333 | 0.490562 | 0.435739 | 0.342108 | 0.302127 | 0.648685 |
| KNN | 0.484520 | 0.495124 | 0.435529 | 0.440416 | 0.430759 | 0.429378 | 0.455146 |
| DT | 0.607748 | 0.604042 | 0.573608 | 0.555048 | 0.545284 | 0.529802 | 0.607006 |
| Linear SVM | 0.654836 | 0.635218 | 0.579643 | 0.433200 | 0.556313 | 0.419900 | 0.663213 |
| RF | 0.525768 | 0.507213 | 0.356400 | 0.382908 | 0.264574 | 0.308144 | 0.534791 |
| SGD | 0.644125 | 0.619413 | 0.592901 | 0.560136 | 0.545816 | 0.537861 | 0.660135 |
| EHC | 0.802871 | 0.713470 | 0.801173 | 0.840199 | 0.727786 | 0.830871 | 0.904512 |

### 4.3.1 Restaurants Review Dataset

*Figure 3* and *figure 4* shows the performance parameter comparison of Restaurant review classification task using Proposed hybrid ensemble classifier with various train – test split ratio. The results of 70% - 30% split ratio is better compared to others in terms of accuracy, precision, recall and F1-score.



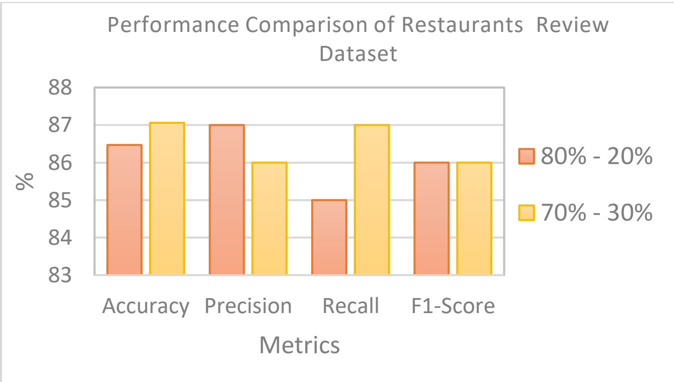


**Figure 3.** Performance Comparison of Restaurant Review Dataset with different size of train - test split

FOREX Publication
Open Access | Rapid and quality publishing

International Journal of
Electrical and Electronics Research (IJEER)
Research Article | Volume 12, Special Issue -BDF | Pages 28-35 | e-ISSN: 2347-470X

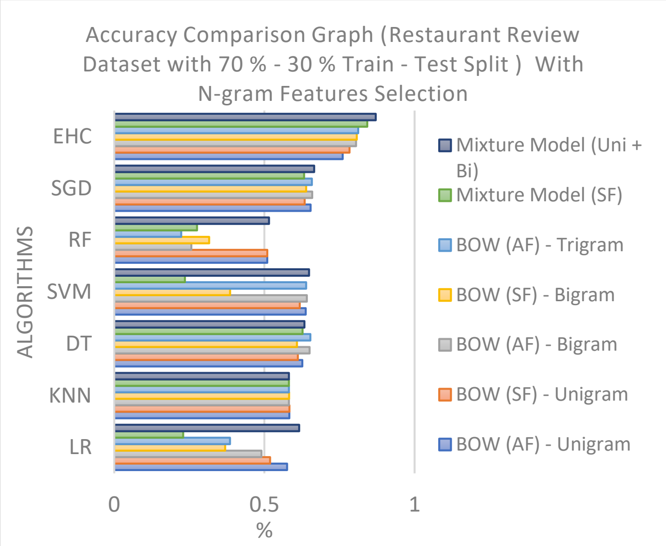


**Figure 4.** Restaurant review dataset accuracy comparison graph (70% - 30% Split)

### 4.3.2 Laptop Review Dataset

*Figure 5* and *figure 6* shows the performance parameter comparison of laptop review classification task using proposed hybrid ensemble classifier with various train – test split ratio. The results of 70 % - 30 % split ratio is better compared to others in terms of accuracy, precision, recall and F1-score.

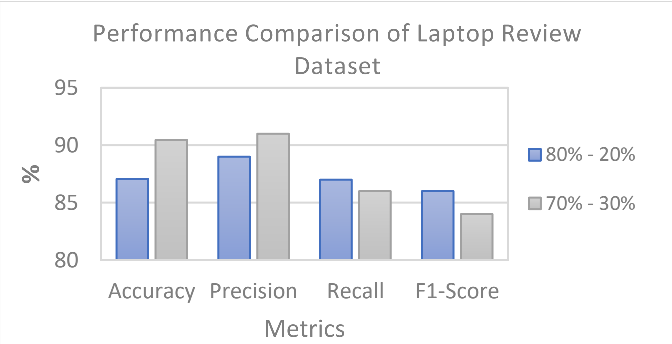


**Figure. 5** Performance Comparison of Laptop Review Dataset with different size of train - test split

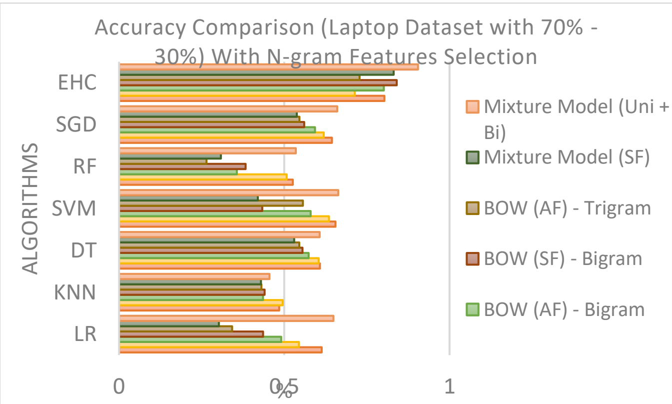


**Figure 6.** Laptop review dataset accuracy comparison graph (70% - 30% Split)

## 5. CONCLUSION

This work focuses on enhancing sentiment analysis and fake review detection within the context of a social media review. Here various challenges that commonly encountered when analyzing raw online reviews are captured and propose a N-gram based features selection technique and hybrid ensemble machine learning classification model as a solution to tackle those problems. This method transforms pre-processed reviews into meaningful feature vectors, enabling efficient, reliable, and robust classification. Our results demonstrate that the hybrid approach significantly improves sentiment analysis and fake review detection performance compared to individual machine learning methodologies. We also compare our model's performance with state-of-the-art methods across multiple evaluation parameters, including F-1 measure, accuracy, all of which shown marked improvement.

## REFERENCES

[1] Balli, Cagla&Guzel, Mehmet &Bostanci, GaziErkan& Mishra, Alok. (2022). Sentimental Analysis of Twitter Users from Turkish Content with Natural Language Processing. Computational Intelligence and Neuroscience. 2022. 10.1155/2022/2455160

[2] Akhtar MS, Gupta D, Ekbal A, Bhattacharyya P. Feature selection and ensemble construction: a two-step method for aspect based sentiment analysis. Knowl Based Syst. 2017; 125:116–35

[3] Fan, Feifan & Feng, Yansong & Zhao, Dongyan. (2018). Multi-grained Attention Network for Aspect-Level Sentiment Classification. 3433-3442. 10.18653/v1/D18-1380.

[4] Shraddha Kalbhor, Dinesh Goyal; Survey on ABSA based on machine learning, deep learning and transfer learning approach. AIP Conf. Proc. 15 June 2023; 2782 (1): 020041. https://doi.org/10.1063/5.0154549

[5] Manek, A.S., Shenoy, P.D., Mohan, M.C. et al. Aspect term extraction for sentiment analysis in large movie reviews using Gini Index feature selection method and SVM classifier. WWW 20, 135–154 (2017).

[6] Pranckevicius, Tomas & Marcinkevičius, Virginijus. (2017). Comparison of Naive Bayes, Random Forest, Decision Tree, Support Vector Machines, and Logistic Regression Classifiers for Text Reviews Classification. Baltic Journal of Modern Computing. 5. 10.22364/bjmc.2017.5.2.05.

[7] Ceron, Andrea & Curini, Luigi &Iacus, Stefano. (2014). Using Sentiment Analysis to Monitor Electoral Campaigns: Method Matters--Evidence from the United States and Italy. Social Science Computer Review. 33. 10.1177/0894439314521983.

[8] L. Yang, Y. Li, J. Wang and R. S. Sherratt, "Sentiment Analysis for E-Commerce Product Reviews in Chinese Based on Sentiment Lexicon and Deep Learning," in IEEE Access, vol. 8, pp. 23522-23530, 2020, doi: 10.1109/ACCESS.2020.2969854

[9] Z. Kastrati, A. S. Imran and A. Kurti, "Weakly Supervised Framework for Aspect-Based Sentiment Analysis on Students‟ Reviews of MOOCs," in IEEE Access, vol. 8, pp. 106799-106810, 2020, doi: 10.1109/ACCESS.2020.3000739.

[10] Lv, Yanxia& Wei, Fangna& Cao, Lihong& Peng, Sancheng&Niu, Jianwei& Yu, Shui& Wang, Cuirong. (2020). Aspect-level Sentiment Analysis Using Context and Aspect Memory Network. Neurocomputing. 428. 10.1016/j.neucom.2020.11.049.

[11] Yan, Wei & Zhou, Lifan& Qian, Zhengjiang& Xiao, Le & Zhu, Haixia. (2021). Sentiment Analysis of Student Texts Using the CNN-BiGRU-AT Model. Scientific Programming. 2021. 10.1155/2021/8405623

[12] M. E. Basiri et al., "Improving Sentiment Polarity Detection Through Target Identification," in IEEE Transactions on Computational Social Systems, vol. 7, no. 1, pp. 113-128, Feb. 2020, doi: 10.1109/TCSS.2019.2951326.

[13] Khetani, Vinit, et al. "Cross-Domain Analysis of ML and DL: Evaluating their Impact in Diverse Domains." International Journal of Intelligent Systems and Applications in Engineering 11.7s (2023): 253-262.

[14] Al-Shammari, Reham&Yousif, Suhad A. (2020). Fake News Classification Using Random Forest and Decision Tree (J48). 23. 8

Open Access | Rapid and quality publishing

# International Journal of Electrical and Electronics Research (IJEER)

Research Article | Volume 12, Special Issue -BDF | Pages 28-35 | e-ISSN: 2347-470X

[15] A. Ishaq, S. Asghar and S. A. Gillani, "Aspect-Based Sentiment Analysis Using a Hybridized Approach Based on CNN and GA," in IEEE Access, vol. 8, pp. 135499- 135512, 2020, doi: 10.1109/ACCESS.2020.3011802.
[16] M. U. Ghani, M. Rafi and M. A. Tahir, "Discriminative Adaptive Sets for Multi-Label Classification," in IEEE, vol. 8, pp. 227579-227595, 2020.
[17] J. Huang, H. Rui, G. Li, X. Qu, T. Tao and X. Zheng, "Multi-Label Learning with Hidden Labels," in IEEE Access, vol. 8, pp. 29667-29676, 2020, doi: 10.1109/ACCESS.2020.2972599.
[18] Deng S, Sinha AP, Zhao H. Adapting sentiment lexicons to domain-specific social media texts. Decis Support Syst. 2017; 94:65–7.
[19] Siering M, Deokar AV, Janze C. Disentangling consumer recommendations: explaining and predicting airline recommendations based on online reviews. Decis Support Syst. 2018; 107:52 63.
[20] Kermani FZ, Eslami E, Sadeghi F. (2019). Global Filter--Wrapper method based on classdependent correlation for text classification. Engineering Applications of Artificial Intelligence, 85, 619-633
[21] Li, L.; Yang, L.; Zeng, Y. Improving Sentiment Classification of Restaurant Reviews with Attention-Based Bi-GRU Neural Network. Symmetry 2021, 13, 1517. https://doi.org/10.3390/sym13081517.