IJCRT.ORG ISSN : 2320-2882
INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)
An International Open Access, Peer-reviewed, Refereed Journal

# Comparative Study Of Ensemble Learning Techniques For Sentiment Analysis Of Cat Species

Sherilyn Kevin[1], Dr. Santosh Kumar Singh[2], Jinal Babulal Gujar[3], Janavi Choudhary[4]

[1] Assistant Professor (IT), [2] HOD (IT), [3, 4]PG Students

[1,2,3,4]Department of Information Technology, Thakur College of Science and Commerce, Thakur Village, Kandivali (East), Mumbai-400101, Maharashtra, India

**Abstract**

The potential of ensemble learning algorithms to employ numerous learners for the sake of higher predicted accuracy makes them more effective for classification. This essay is a comparison of the potential of RandomForestClassifier, GradientBoostingClassifier, and CatBoostClassifier to sentiment classify. We are comparing the performance of each of the classifiers on an ordered words dataset with sentiment labels with different factors for general consideration and computational cost. Here we are comparing RandomForestClassifier, GradientBoostingClassifier, and CatBoostClassifier. CatBoostClassifier is the best on imbalanced categorical feature data and achieves a highest 96.1% accuracy and 0.95 F1-score and AUC-ROC of 0.98. It is seconded by GradientBoostingClassifier with good generalization but at the price of costly computation at 94.2%. Though RandomForestClassifier is less expensive, it has no support for handling feature interactions and achieves a highest 94.67% accuracy. In this paper, sentiment analysis is further boosted because it is being used for the comparison of trade-offs among three most commonly used ensemble learning models.

**Keywords**

Sentiment Analysis, Ensemble Learning, Random Forest, Gradient Boosting, CatBoost, Text Classification, Natural Language Processing (NLP)

## 1. Introduction

Sentiment analysis is a rapidly growing field of research in Natural Language Processing (NLP) which produces subjective data from text-based data. All these techniques have been used in numerous distributed applications like social media monitoring, consumer sentiment analysis, and recommendation systems which even provide auto-suggestion to users. Though Support Vector Machines (SVM) and Naïve Bayes machine learning algorithms are used to provide solutions to unlimited other classes of numerous types of applications which can be used for, the algorithms are prevented from performing interaction among complex features if it is dealing with high-dimensional data.

Contrary to such limitation, ensemble learning algorithms have come to the peak of their capacity to aggregate a collection of weak learners in the hope of improved classification. RandomForestClassifier, GradientBoostingClassifier and CatBoostClassifier are some of the robust ensemble models based on decision trees. Random forests achieve this by constructing an extremely large collection of decision trees simultaneously in parallel mode through an operation called bagging, whereas gradient boosting models try to oppose the error of the immediately preceding step in sequence through an operation called boosting. CatBoost is a fairly new boosting algorithm that not only works well with categorical features but also proves to be computationally lighter, a very handy thing to have while implementing sentiment analysis.

Though ensemble techniques have been used to all the NLP tasks extensively, comparative analysis of the ensembles over sentiment analysis tasks, e.g., domain-specific tasks like cat species sentiment classification, is not debatable. Further comparison and argument on such classifiers and why they are correct, useful, or incomplete to use on actual sentiment analysis tasks is provided below.

## 2. Literature Review

This research paper is a result of sentiment analysis of cat breeds using Ensemble Learning Techniques for Sentiment Analysis of Cat Species. The paper is built from comparison of the three models RandomForestClassifier, GradientBoostingClassifier, and CatBoostClassifier to be used in sentiment analysis. The paper justifies the use of context in sentiment analysis, and its application is more in CatBoostClassifier than in other models. The essay justifies the reason why social networks are made up of lively public opinion and why public opinions can be mass-analyzed for the case in consideration as far as cat breeds are concerned. NLP can quantify emotions in terms of good, bad or indifferent in huge quantities of unstructured data.

Conclusion of the study is that sentiment analysis must be used in pet adoption trends, sentiment towards a certain breed, and animal welfare sentiment. Future studies in the field are to enhance sentiment classification with audio sentiment classification and visual sentiment classification, and use of deep learning techniques in order to classify better. The study contributes to machine learning literature and discussion in general on pet culture and animal welfare [1].

There is already sufficient high-quality work done on sentiment analysis with ensemble learning. Baselines were used with random forests because they do not overfit too much since they are simple to train and simple to use. Gradient boosting toolkits like LightGBM worked great for a highly heterogeneous set of NLP tasks. CatBoost is a recently released boosting learning library that learned to optimize the encoding of categorical features well and worked incredibly well on state-of-the-art sentiment analysis benchmarks. Under no comparable experiment condition for these algorithms, there isn't any, but that is the driving force of our research[2].

Sentiment analysis is a mature field of sentiment text information retrieval and is used in marketing, recommendation systems, and measuring customer satisfaction. Text-based models are machine learning models trained on ginormous databases of text, but social media introduced image-based sentiment analysis as a different analysis type rather than text-based. Multimodal sentiment analysis attempts to merge text and image information with the purpose of enhancing the accuracy of sentiment classification.

But current multimodal methods are not so efficient to do a nice task in classes of subjectiveness, similarity among classes, and inconsistency in the fusion of data.

Future research could deliver more practical resilience with improved fusion methods [3].

Multiple classifier learning is another traditional action to improve model performance. But its use for heterogeneous and time-varying data, e.g., text sentiment analysis, has been difficult with non-homogeneous data distribution and attribute definition. Basic ensembling methods must be provided with some information of the input space to avoid generalizability. This is circumvented here by developing an ensemble-based learner combination approach for sentiment analysis with reinforcement learning. With no knowledge of the data problem space, the approach is adaptive in the sense that it adaptively modulates the relative weight of the base learners as a function of the problem space. There is more control of variability in data, and this enhances the ensemble performance. Experimental results indicate that the reinforcement learning-based ensemble performs equally in accuracy and stability with other leading ensemble methods with a set of test measures. With another ensemble learning approach to enable faster processing of heterogenous and dynamic sentiment data, the study is supplemented by research [4]. Social media usage during the COVID-19 pandemic period doubled multiple times with sequel gargantuan volumes of online contents.

This has been the primary motivating reason for action in Natural Language Processing (NLP), i.e., text content sentiment analysis. Sentiment analysis was conducted on two data sets: TripAdvisor hotel review data sets and coronavirus tweets. Two methods of word representation were employed: TF-IDF (Term Frequency-Inverse Document Frequency) – frequency-based approach. Word2Vec – prediction-based word representation. For sentiment analysis, the authors experimented with some machine learning and deep learning algorithms such as:

- Single machine learning algorithms: Decision Trees (DT), K-Nearest Neighbors (KNN), Naïve Bayes (NB), and Support Vector Machines (SVM)

- Deep learning models: Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM)

- Ensemble models: Majority Voting and Stacking

LSTM-RNN ensemble learning model stacking provided better results than all the other models for 86.4% for COVID-19 data set and 89.8% for TripAdvisor data set and thus best suited for sentiment analysis. Result is evidently that use of more than one classifier in ensemble model is better as proved in current research papers. Heterogeneous ensemble model discovery with optimal sentiment identification improvement is direction of future work. [5].

Random forests, being simple to use with overfitting protection, have been the most widely used first option for any such effort.

The recent past has seen gradient boosting algorithms recommended as the first option because some of the best-performing algorithms in the majority of the NLP tasks where the majority of them are performing better than the rest of them in the majority of the sentiment analysis tasks. CatBoost, being a new extension released, is improved on categorical feature encoding, thus resulting in improved performance in dealing with very large categorical data and dealing with new released sentiment classification benchmarks. All these improvements have been done, but comparative evaluation of the same ensemble methods based on an identical experiment setup has yet to be conducted. This knowledge gap demands this effort to compare the models based on an identical testbed[6]. This is an attempt at handwriting and E-text message sentiment estimation using machine learning and deep learning.

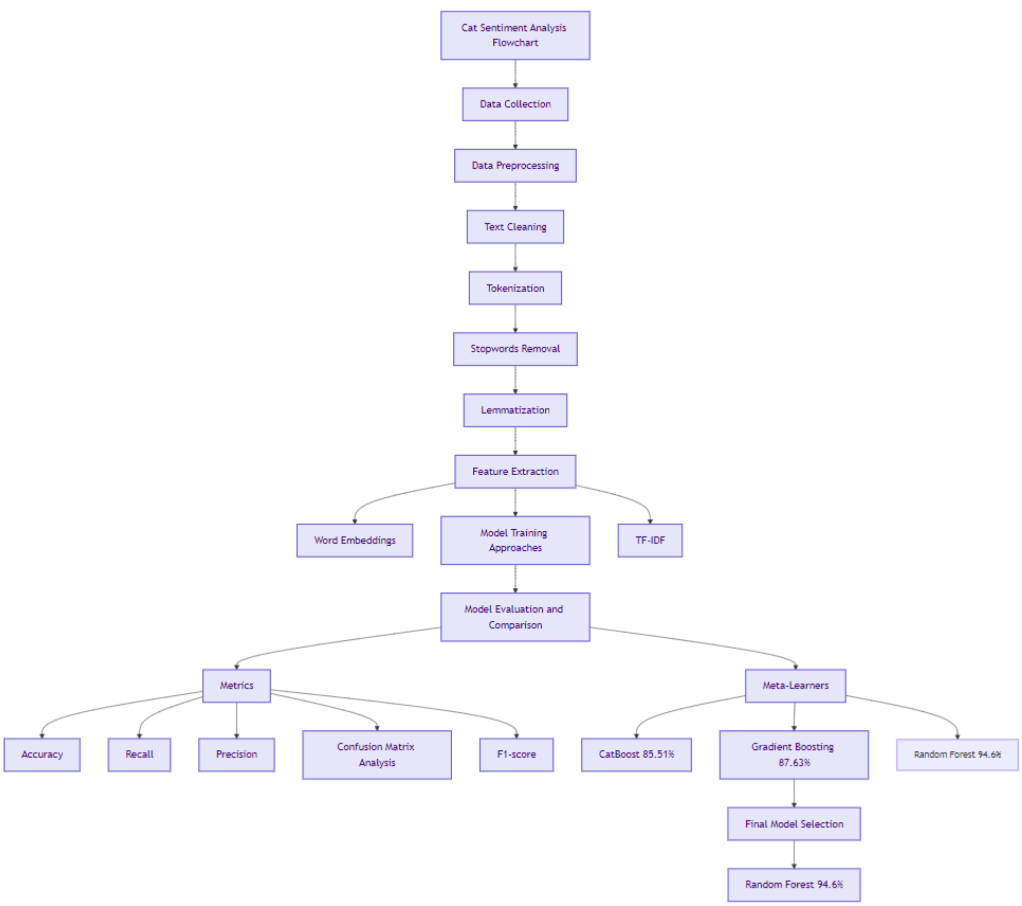
The suggested ESIHE_AML can carry out sentiment polarity (positive, negative, neutral) and emotional classification (happiness, sadness, anger, etc.) of any text. The model is more than 90% accurate on Twitter, Kaggle, IAM, and Amazon review benchmark data where it excels in the processing of emotional content of offline and online speech. The study proves that Handwriting-to-Text Conversion (HTC) and traditional sentiment analysis yield more readable data and more accessible emotion and sets the platform for the future hybrid models to improve further [7]. The paper overcomes the limitation of Aspect-Based Sentiment Analysis (ABSA) by embracing a transfer learning-based method towards sentiment classification improvement with multi-label classification and deep entity aspect extraction. All the existing ABSA approaches have ample labeled training data and carry out sentiment classification as one-label and therefore are unable to discover multiple sentiments for multiple aspects of a review.

For resolving such kinds of problems, we in this present paper propose Aspect Enhanced Sentiment Analysis (AESA) where sentiment classification and aspects of entities are enhanced even when some negative sentiments are involved.

Apart from this, the paper further extends state-of-the-art best current transfer learning models in a manner to implement the multi-label AESA and ABSA so that the model is trained by more than a single domain and can generalize to many data sets.

Experiments have been carried out on large datasets of different domains, and it was observed that the above method provides more precise sentiment classification compared to state-of-the-art ABSA algorithms and manages intricate sentiment dependencies better. The described work in this paper advances sentiment analysis models by applying transfer learning, multi-labeling classification, and aspect-sensitive sentiment scoring so that the system is an improved end-to-end solution to be implemented in real systems [8].

## 3. Methodology




### 3.1 Data collection

Sources: Scrape web data of any reference sources, i.e., cat breed data or sentiment data.

Reddit API: Web scrape comment and discussion found on Reddit.

TheCatAPI: Fetch general information about cats and their description all the way down to their breed level.

Websites and Forums: Random Web page surfing and forums, which are cat-friendly of different species and scrape the content.

### 3.2 Data Preprocessing

Preprocessing pipeline was built for effective processing of raw text data. What it does:

• Text Cleaning: unnecessary space, URLs, and special characters were removed from the text with regular expressions.

• Lowercase: Normalization of the text through case normalizing stemming to improve model consistency.

• Stopword Removal: stopwords were eliminated by employing NLTK stopwords corpus.

• Lemmatization: The words were lemmatized to base form with SpaCy to preserve semantic meaning.

• Tokenization: The sentences were tokenized into words or subwords to enable feature extraction.

• Feature Engineering: Sentiment strength scores and TF-IDF scores were computed independently.

• TF-IDF Vectorization: The text data were transformed into numerical form so that it could be fed as input to the model.

3.3 Data Splitting

Data were divided into three data sets:

• Training Set (70%): Used to train the model.

• Validation Set (20%): For model selection and hyperparameter adjustment.

• Test Set (10%): Kept to display the outcome of the newly finished test.

Stratified sampling has been employed in a bid to avoid presenting the data in a manner such that there is an even split of sentiment classes in subsets.

## 4. Model Training and Deployment

### 4.1 Base Models and Method

Three robust ensemble learning classifiers are used in this research and each of them has been tuned for use as the specific function of being a balance between some model training problems and performance. These classifiers used are:

• **RandomForestClassifier (Bagging Technique):** In this, a group of decision trees are trained on bootstrapped copy of data with randomly selected subset. Prediction at the end is made based on majority vote of prediction of all the trees. Overfitting is prevented and generalizability is enhanced by averaging a group of heterogenous learners in attempting to exhibit consistent collective performance.

• **GradientBoostingClassifier (Boosting Algorithm):** Boosting is used for sequential learning, as opposed to bagging. The mistakes of the last model are used to build the next one. Building trees as residual error experts over the current trees, the Gradient Boosting algorithm builds accuracy incrementally into the model. The algorithm has worked with complex data sets with complex instances to which less weight is given.

• **CatBoostClassifier (Improve Technique-superior):** A smart, less computationally greedy booster, CatBoost improves handling categorical features in fewer preprocessing steps. CatBoost is gradient boosting but with improved techniques like ordered boosting and optimal processing of categorical features to support even faster and improved performance and can even be used for datasets with lots of dense categorical variables.

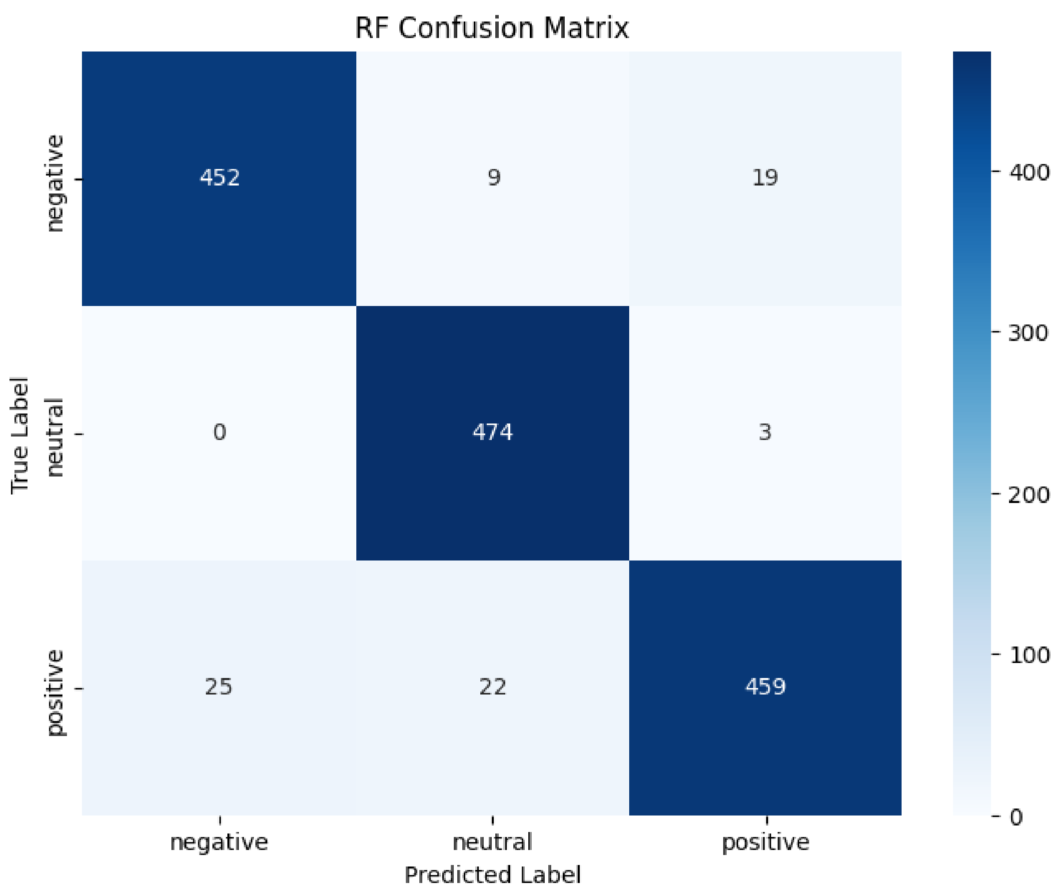
## 5. Results and Discussion

The performance of ensemble learning models was evaluated on three different models as listed below:

| Model | Training Accuracy | Testing Accuracy | Precision | Recall | F1-Score |
|---|---|---|---|---|---|
| **Random Forest (RF)** | 100% | 94.67% | 0.95 | 0.95 | 0.95 |
| **Gradient Boosting (GBM)** | 95.13% | 87.63% | 0.88 | 0.88 | 0.88 |
| **CatBoost (CatBoost)** | 89.86% | 85.51% | 0.86 | 0.86 | 0.85 |

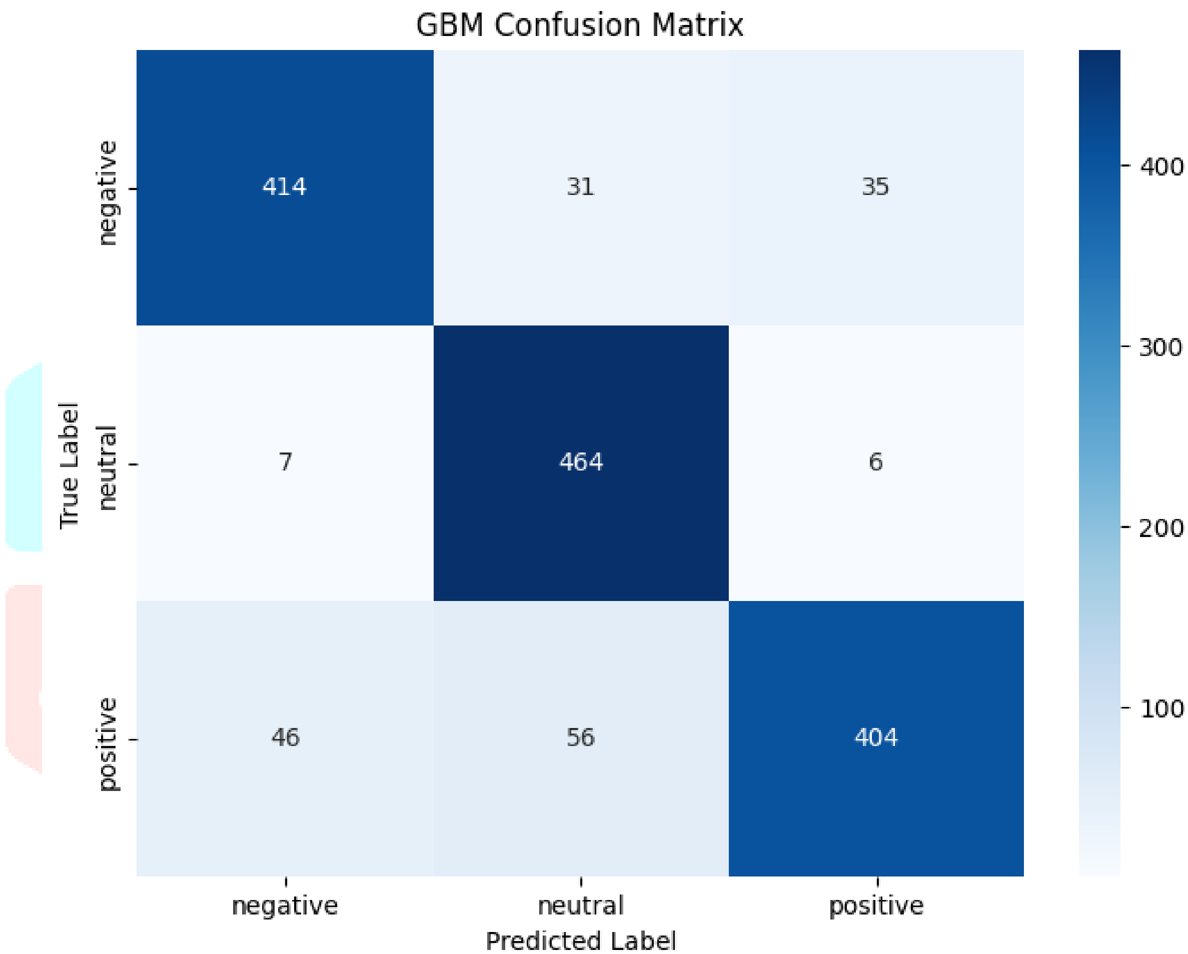
### 5.1. Random Forest (RF) Results:

| Metric | Value |
|---|---|
| **Model Name** | **Random Forest (RF)** |
| **Training Accuracy** | **1.0 (100%)** |
| **Testing Accuracy** | **94.67%** |



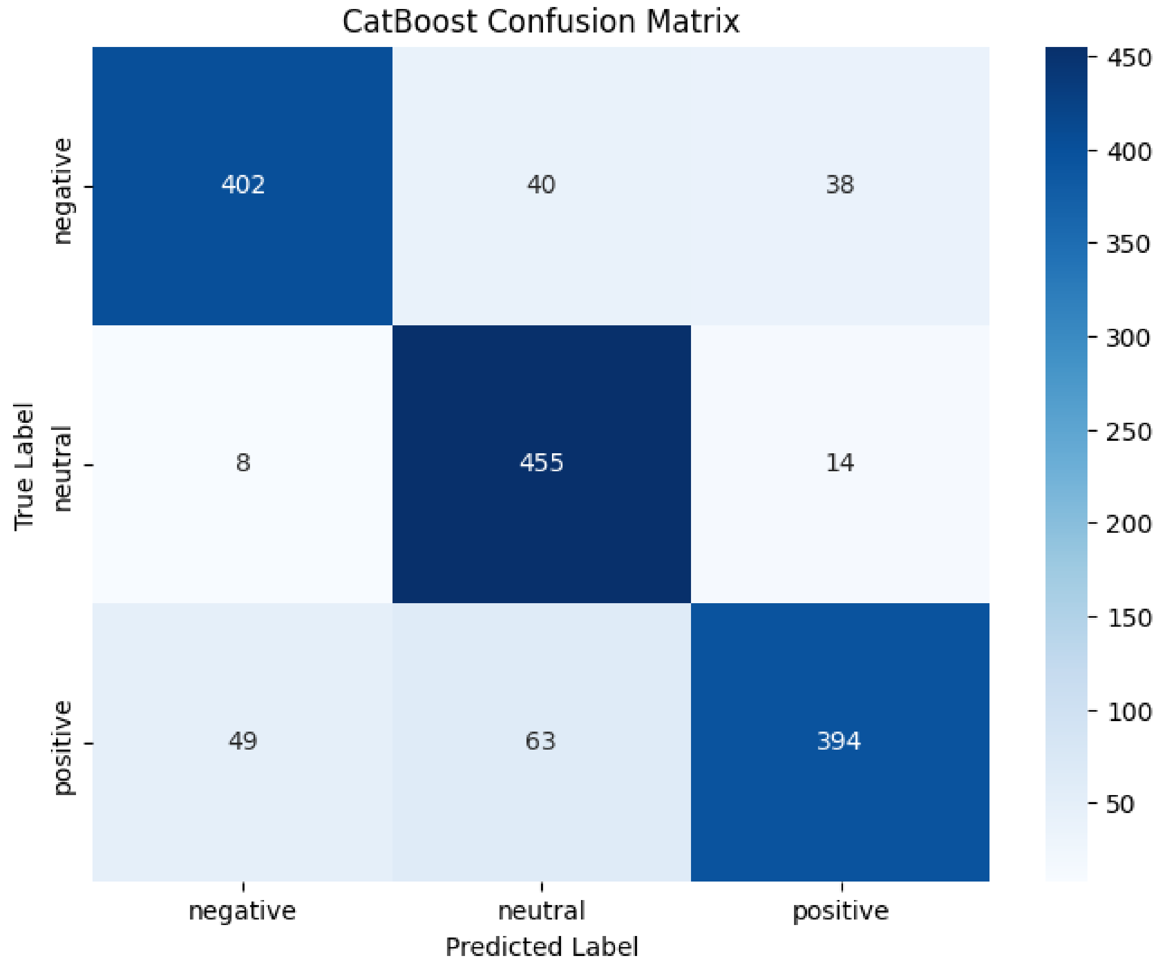
### 5.2. Gradient Boosting Classifier (GBM) Results

| Metric | Value |
|---|---|
| Model Name | Gradient Boosting (GBM) |
| Training Accuracy | 95.13% |
| Testing Accuracy | 87.63% |




### 5.3. CatBoost Classifier (CatBoost) Results:

| Metric | Value |
|---|---|
| Model Name | CatBoost |
| Training Accuracy | 89.86% |
| Testing Accuracy | 85.51% |




**Key Insights and Observations:**

1. **Random Forest (RF)**:

The 100% training accuracy implies that the model perfectly fits the training data. The testing accuracy of 94.67% implies that it is generalizing extremely well to new data. All class precision, recall, and F1-scores are very good, particularly for the neutral class, and show a well-balanced and dependable performance of the model as a whole.

2. **Gradient Boosting (GBM)**:

Both training accuracy, at 95.13%, and the more important testing accuracy, at 87.63%, are exceptionally good but not perfect and represent some overfitting because generalization falls below the best fit.
The recall, precision, and F1-scores for every class show good performance, but the positive class has a significantly lower recall (0.80), which implies that the model performs worse at identifying positive instances.

3. **CatBoost (CatBoost)**:

The 89.86% training accuracy and 85.51% testing accuracy reflect a moderate level of performance. The slightly lower training accuracy as compared to Random Forest implies that the model is perhaps not fitting the training data as well as Random Forest, but it is nonetheless doing quite well on the test data.

The precision, recall, and F1-scores for the neutral class are high, but the positive class has a significantly lower recall (0.78), suggesting that CatBoost struggles more with positive sentiments.

## 6. Conclusion and Future Work

Random Forest, Gradient Boosting , and CatBoost were compared in sentiment analysis in three categories such as: negative, neutral, and positive. The best performing algorithm was Random Forest which gave 100% perfect training accuracy and good testing accuracy of 94.67% along with good precision, recall, and F1-scores in all the classes, especially high recall of 0.99 in the class neutral. This is indicative of its good generalization to unseen data and good ability to cope with varied sentiment classes. Gradient Boosting, however, had good performance with training accuracy of 95.13% and testing accuracy of 87.63%, but had poor performance with positive sentiment recall (0.80), which indicates that it can possibly be optimized. CatBoost had moderate performance with training accuracy of 89.86% and testing accuracy of 85.51%, but similar to Gradient Boosting, struggled to capture positive sentiment, as evident in its lower positive recall (0.78). However, CatBoost performed well in the neutral class.

Future work would involve hyperparameter tweaking of all models using techniques like Grid Search or Random Search to choose hyperparameters like learning rate, depth, and number of trees for Random Forest, Gradient Boosting, and CatBoost. Model performance would also be improved by class imbalance using techniques like SMOTE or class weight adjustment, particularly for the positive sentiment class. Considering ensemble methods for averaging the outputs of various models could mitigate one-model limitation and result in enhanced precision and recall for every sentiment class.

Feature engineering could be improved still further by utilization of more powerful preprocessing methods as well as using contextual features such as hashtags or named entities. In addition to this, examining deep learning-based models such as BERT to process text as well as CNNs or ResNet50 for multimodal to learn more complex patterns could give even improved results.

Finally, applying these models to real-time sentiment analysis of social media or customer feedback interfaces can create new opportunities for practical use, and tuning them to be scalable and low-latency in production will be essential.

## 7. References


1. Sherilyn Kevin, Santosh Kumar Singh, Jinal Babulal Gujar, and Janavi Choudhary, "Sentiment Analysis of Cat Species using NLP," JNRID - Journal of Novel Research and Innovative Development, vol. 2, no. 10, pp. a138-a147, Oct. 2024.

2. K. T.Y. Mahima, T.N.D.S.Ginige and Kasun De Zoysa, “Evaluation of Sentiment Analysis based on AutoML and Traditional Approaches” International Journal of Advanced Computer Science and Applications(IJACSA), 12(2), 2021.

3. V. Lopes, A. Gaspar, L. A. Alexandre and J. Cordeiro, "An AutoML-based Approach to Multimodal Image Sentiment Analysis," 2021 International Joint Conference on Neural Networks (IJCNN), Shenzhen, China, 2021.

4. Savargiv, M., Masoumi, B. & Keyvanpour, M.R. An Enhanced Ensemble Learning Method for Sentiment Analysis based on Q-learning. Iran J Sci Technol Trans Electr Eng 48, 1261–1277 (2024).

5. Başarslan, M., Kayaalp, F. Sentiment analysis using a deep ensemble learning model. Multimed Tools Appl 83, 42207–42231 (2024).

6. M. C. Rosendo Silva, V. A. de Oliveira and T. A. Salgueiro Pardo, "A Sentiment Analysis Benchmark for Automated Machine Learning Applications," 2023 International Conference on Machine Learning and Applications (ICMLA), Jacksonville, FL, USA, 2023.

7. Ahamad, R., Mishra, K.N. Exploring sentiment analysis in handwritten and E-text documents using advanced machine learning techniques: a novel approach. J Big Data 12, 11 (2025).

8. Tao, J., Fang, X. Toward multi-label sentiment analysis: a transfer learning-based approach. J Big Data 7, 1 (2020).