Pakistan Journal of Engineering and Technology, PakJET
Multidisciplinary & Peer Reviewed
Volume: 4, Number: 2, Pages: 67- 75, Year: 2021

A Hybrid Approach for Feature Extraction and Classification Using Machine Learning Techniques

Asim Alvi, Kashif Sultan, and Adeel Muzaffar Syed

Software Engineering Department, Bahria University, Islamabad, Pakistan
Corresponding author: Asim Alvi (Asimalvi581@gmail.com)

Abstract- **Artificial intelligence (AI) solutions are used to help make choices that include a high precision of choices they recommend and a deep understanding of choices, so that the chiefs can trust them. Verifiable, non-emblematic learning methods have greater perceptive accuracy. Express, fair representation information learning method produces increasingly justifiable Approach's. Hybrid AI systems analyses the data and exploratory characteristics of approach types. The fundamental purpose of this commitment is to differentiate between a proper AI strategy for choice of assistance that produces reliable and fair outcomes, depending on the various AI techniques, which provide an analysis of different approaches either with special or similar datasets. Comprehensibility is measured subjectively, however, by the form of learning process and the scale of the resulting representation of knowledge. We need a hybrid approach for classification using feature extraction in preprocessing of imbalanced dataset (raw dataset) by normalization for better classification, accuracy, and minimization of Error Rate for the smaller amount of training of significant dataset and lesser amount of training time for a good result. Moreover, data preprocessing play very important role in the field of machine learning, the result is more accurate if your data is cleaned otherwise some misclassification occurs which tend to be problematic in classification problems. Therefore, hybrid comparative analysis and approach is required for the selection of dataset with respect to machine learning classifier that will having different results with different data sets based on the hybrid approach to achieve maximum best results against any dataset. Input with maximum accurate result will be reproduced from our hybrid approach because this approach shows which type of classifiers should be used under what type of dataset you have, meanwhile exception of the generous fact based on results should be different among different classifiers when applied to different dataset. After that, a comparative analysis of different algorithms with different dataset has made and a comparison shows the enhanced significant generic approach with the hybrid approach, a clear result in the form of accuracy, precision, recall and f1 score shows the results against the specific techniques against the dataset and that tell us about the accuracy and rate of misclassification error. At the end we will see which machine learning algorithm improves the accuracy for which type of dataset.**

***Index Terms*—Hybrid Model, Machine Learning Techniques, Deep learning, Misclassification, Classification, Feature Extraction, Neural Networks.**

I. INTRODUCTION

Computational effectiveness in displaying and anticipating is a certain degree of leeway over certain other arrangement calculations, which is possible due to the fact of simple parallelization, particularly useful for massive data [1]. The Naive Bayes classifier is defined as computational competence, low fluctuation, constant learning, and direct prediction of back probability, clamor strength and resilience of missing characteristics [2]. Text characterization is a large and fundamental undertaking in controlled Artificial Intelligence (AI). Its application is in the field of email spam recognition, slant inquiry, language area, and character development and so on. Various classifiers can be used to characterize the archive. Many of them are neural networks, support vector machines, genetic equations, Naive Bayes classifiers, k-Nearest Neighbor (kNN) [3]. The key issues in creating high-profile determination processes are the collection of a small list of capabilities to reduce the cost and running time of the framework, as well as the achievement of an acceptably high recognition rate [4]. In recent years, a variety of post-pruning estimates have been provided, such as reduced blunder pruning, error-free pruning, less blunder pruning, and cost-based pruning. For example, most of the culling methodologies decreased the error culling and the least blunder pruning of the selected tree in the base on request evaluating the misclassifying blunders for each over fitting [5].The expected bit trick encourages us to make such an improvement inside the SVM, despite having to take care of a huge penalty as far as computational skills are concern [6]. The degree to which a person expresses his or her expectations is limited on a regular basis when individuals need to give assessments on an item in the form of score / star assessments. In any case, whenever an individual is allowed to interact audits in the sort of system interfaces, he can be highly accurate as to what dimensions to the object are appropriate and what is certainly not normal [7].

Colon disease is one of the most well-known dangerous tumors with a high frequency in the age gathering of 40-50 years and is an intense danger to human life and wellbeing. In 2017, around 136,000 new instances of colon malignant growth were analyzed. Around 1 out of 20 individuals will create colon malignant growth during their lifetime [8]. Since the dataset takes 127 Metallica and 80 Nirvana tunes, the Naive Bayes Classifier has been castoff on the grounds that it is reasonable to use a small dataset [9]. As a result, the proposed equation can be organized as solid, modifying each of the 28 images would generate new concept capacities. It managed to bring about the capacity of the premises. The development of the ICA premises models is conducted following the application of a large-scale reduction of observations as provided by the PCA using the Naïve Bayes Classifier [10].

A new estimate using an innocent Bayes seeing the on bosom disease expectations has also been developed. The author used details on breast cancer screening for the predictions of the tumor. GUI was designed to discover the possibility of malignant growth in the bosom in women later on. They used a total of nine attributes for prediction. Another parameter choice technique was created by bundling and ordering. In this research, the author focuses with model-based learning, as well as the authors compare and other time - varying-choice approaches on the underway dataset [11]. Bokde et al. have transmitted the R code for showcase. Due to the roughly similar thought experiment making plans of the PSF, the measurement of the k-nearest neighbor (kNN) is an elderly thought experiment machine and is flawlessly updated [12]. It is often used to deal with non-linear problems, such as Credit score and bank customer rankings, in which the information collected does not necessarily obey the conventional straight assumption, and it should be one of the first decisions when there is almost zero earlier knowledge on distribution information. In addition, the impact of the variables on the test types can be effectively minimized [13]. kNN has sophisticated predicting accurateness and takes no conventions for the composed data, and mainly, it is not penetrating to the outliers. It has been extensively practical in actual-ecosphere glitches, like investigating the area lie under the stock market, so that daily up and down of shares in the market [14]. KNN-grounded classification accuracy might expressively be contingent on the rules that are executed to compute reserves between training and testing samples [15].

Towards discourse this subject, the biased reserve rules and the depiction-based approaches are used to calculate the resemblances of samples to regulate k-nearest neighbors of apiece demanded mock-up [16].

kNN strategy is such a well-known information extraction and insight classification technique due to its easy application and impressive arrangement execution. It is, however, irrational for traditional kNN approaches to demote fixed K confidence (regardless of the fact that it has been developed by experts) to all tests [17]. Tree structures where the subjective parameter may take the shape of a given configuration of attributes are called order trees, In this tree structure, the leaf relate to the identities of the class and the nodes speak to the conjunctures of the highlights which lead to those class marks which are opposed to the ones used to choose the yield for the classification shown in Tab. I [18].

TABLE I: Dataset Details

Sr. No	Dataset	Details of Dataset	Train/Test Ratio
1	Amazon Mobile Reviews	400,000 Records	70/30
2	Breast Cancer Dataset	56,000 Records	70/30
3	Amazon Alexa Products	34,000 Records	70/30
4	Bi-Language Dataset	40,000 Records	70/30
5	Drugs Record Dataset	161297 Records	70/30

Authors contributed the research work in the following ways;

1. Selection of machine learning classifiers for better results of classification against dataset.
2. Shows which approach and technique will improve the results if you don't know the details regarding dataset.
3. Hybrid Approach working against different datasets to see the accuracy and misclassification rate also the comparative analysis of results in terms of precision, recall, accuracy and f1 score.
4. Shows a significant impact of data filtering techniques against a classification problem.
5. Hybrid Approach Using deep learning and machine learning techniques for better accuracy.

II. RELATED WORK

In this section, we will discuss the literature and find out the importance of machine learning techniques and different classification algorithms based on different datasets and their applications in computing field.

KNN strategy is such a well-known information extraction and insight classification technique due to its easy application and impressive arrangement execution. It is, however, irrational for traditional KNN approaches to demote fixed K confidence (regardless of the fact that it has been developed by experts) to all tests [17]. Decision Tree learning customs a classification tree in place of an insightful technique that plots expectations of a thing to judgments about the value of an object. It is one of the perceptive showcase methods used in measurement, information mining, and AI. Tree structures where the subjective parameter may take the shape of a given configuration of attributes are called order trees, In this tree structure, the leaf relate to the identities of the class and the nodes speak to the conjunctures of the highlights which lead to those class marks which are opposed to the ones used to choose the yield for the classification [18].

In Breiman's technique, the tree in the assortment is formed by choosing a small selection of information (additionally referred to as highlights or factors in the future) to be part of and, respectively, by selecting the best part that depends on these landmarks in the preparation package. The tree is formed using the most severe CART technique, without trimming. This dimensionality random sampling plot is mixed with sacking to reassemble, with partial replacement, the preparation of the instance shows almost every time another tree is formed [19].

In [21] the framework for the analysis of different methods, i.e. the random forests, the exceptional learning machine, the convolutional neural network and the boost vector machine, is implemented in attempt to discover the most capable one. Random forest has been shown to surpass similar classification model in terms of accuracy, steadiness and warmth of recognition, particularly with a little sequence of planning and preparation. As well, contrasting and traditional approaches, random forest are not easily influenced by ecological clamor.

The biggest problem with re-sampling techniques is that, on the basis of few cases, more knowledge analyzing or breaking up will lead to substantial progress. To catch up with this extra vulnerability, Isakson et al. inform the use of certainty span, guaranteeing that the re-sampling of evaluations is unreliable. In either case, Kohavi found both the bootstrap and the cross-approval of decently calculated standard knowledge collections and suggested 10-fold cross-approval as an acceptable accuracy estimation method [22].

SVM is a machine learning classification technique which implements the supervised model for learning, and it is widely used for cancer diagnosis and prognosis field. The SVM technique isolates two classes by choosing a normal classifier that expands the edge. This partition is referred to as the ideal hyper plane insulation. Regularization demarcation and bit work are the two important sections which need to be addressed before guiding the preparation process. A portion of the critical investigations used by SVM to detect the bosom disease used Heuristic analysis SVM approaches, for instance, smooth SVM, wide SVM and general highly nonlinear SVM [24].

Sentiment analysis investigation engines parse through these literary survey data and generate yields in the type of polarity, e.g. positive, negative or neutral. These aids in the determination of the purposes of the important variances in the dealings of the goods and they can be rectified accordingly. Computations in orders may have an impact on the accuracy of the direct consequence in the extreme point and, consequently, on the insufficiency of the grouping [7].

In the incident that accuracy is most noticeable to us, at that moment we should be leaning towards a classification models similar to the Random Forest that manipulate strong learning time nevertheless has the finest accurateness. Mostly on off chance that the preparation of control and cognition is a matter, the Naïve Bayes classifier would be nominated at that moment for its low memory and the need for power handling. On the off possibility that little time to prepare is available but you have the wonderful managing structure and memory then peak Entropy tends to end up being a commendable alternative [25].

This achievement develops more blunder-free packed aside from DT by encouraging the normalization process by ejecting weak apprentices and accelerating Hybrid-target work, the discoveries may not be directly appropriate for different types of AI calculations in anticipation and character development as the unique half-and - half-model may contain various credits for evaluation; In addition, the information arrangement was directed to consider 13 boundaries, such as info and four calm, tropical, cold and hot-bone-dry atmospheres. This calls for further review by endorsing the model for different settings and using larger examples covering numerous structural vitality limits and atmospheres [25].

The fundamental work of this paper shows the order of data for emotional scriptures, that is twitter and item appraisals. The data is separated hooked on as positive and negative. Techniques like Support Vector Machines (SVM), Conditional Random Filed (CRF) and Naive Bayes Multinomial (NBM) remain the calculation prediction based on supervised learning. Three classifiers are individually designed by separating, specific capabilities on basis of features [26].

The SVM classifier, prepared on an unprovoked dataset, will build flawed models that are yet another-sided against the large class and have low execution on the smaller class, as a large part of all the other classification model [27].

A clear multi-name classification technique is dual significance (BR) approach, which breaks the issue down into a series of singular-name multiclass classification issues. A lot of multi-class classifiers are described along these lines which was used to make assumptions. This simple technique, however, it may be, absolutely rejects the conditions between different names. Virtually speaking, multiple publications in a pattern, (For instance, a picture) can have solid relationships or requirements. For example, if a ship classification is shown in a photo, the water class is almost certainly the same in that picture. Misappropriating such trademark dependency could significantly improve the implementation of standards for a multi-name classification [28].

In [29] by evaluating the efficiency of different popular sentiment classification methods as well as developing a comprehensive method that further enhances sentiment classification accuracy. Slight effort has been done in the arena of twitter sentiment analysis of air carriers. This previous work contrasts a variety of different conventional classification techniques and chooses the most specific modelling approach for the implementation of sentiment classification. However, the whole approach that we express increases performance by integrating these sentiment classifiers. In the field of air carriers, the precision of the sentiment classification is high enough just to investigate customer satisfaction. This approach refers to the study of Twitter data by air carriers about their quality of services.

III. METHODOLOGY

The experiment was conducted in two phases. In first phase of experiment we identified the different Datasets and Classifiers also ruined on generic Approach with some data preprocessing techniques for sample dataset to check the novelty. In second phase, based on the filtered dataset received from the phase 1 from original dataset, we created initial step by step experiments and then we carried out the all experiments and recorded them accordingly with the hybrid approach to achieve the maximum best results shown in Fig. 1.

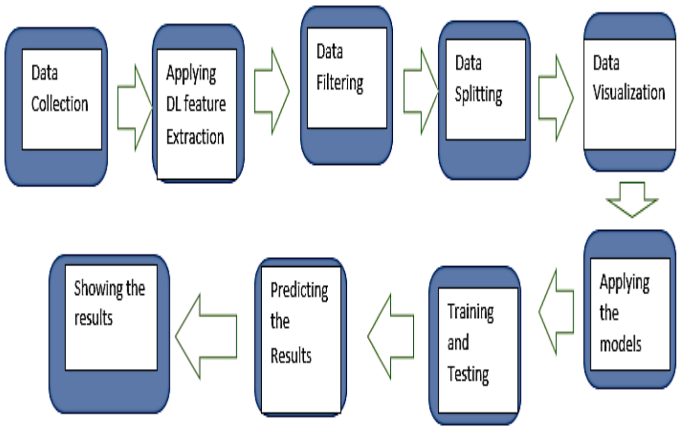


FIGURE 1: Machine learning process

We carried out an experiment on Multiple Datasets. Our experiment is comprised of the several steps.

1. Data Collection (Dataset gathering)
2. Applying Deep Learning for data
3. Data Pre-Processing. (Removing unwanted data or Data cleaning for the features)
4. Identify the novelty of dataset
5. Preparing the dataset according to requirements.
6. Assigning classifiers for checking the performance and behavior on dataset
7. Record the results
8. Visualize the Results for better understanding.
9. Compare the results of different Approach in metrics
10. Mark the best accurate results as leading.
11. Analysis on the experiments with existing knowledge-based results

IV. PROPOSED FRAMEWORK

Existing machine learning approaches have a slightly higher misclassification rate. The accuracy of the machine learning model is degraded due to the higher classification rate. By utilizing a hybrid approach consisting of Machine Learning (ML) and Deep Learning (DL) techniques can reduce the error rates in terms of misclassification of data shown in Fig. 2 and 3.

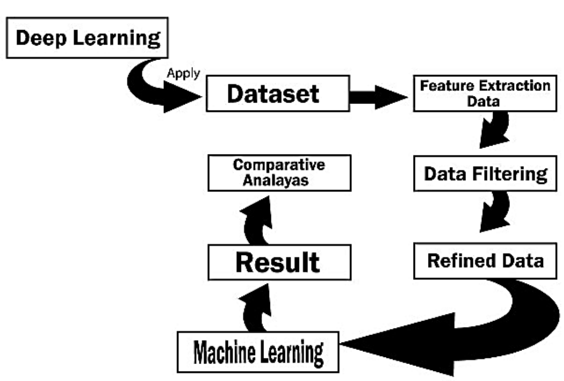


FIGURE 2: Hybrid Approach (Deep learning+ Machine learning process)

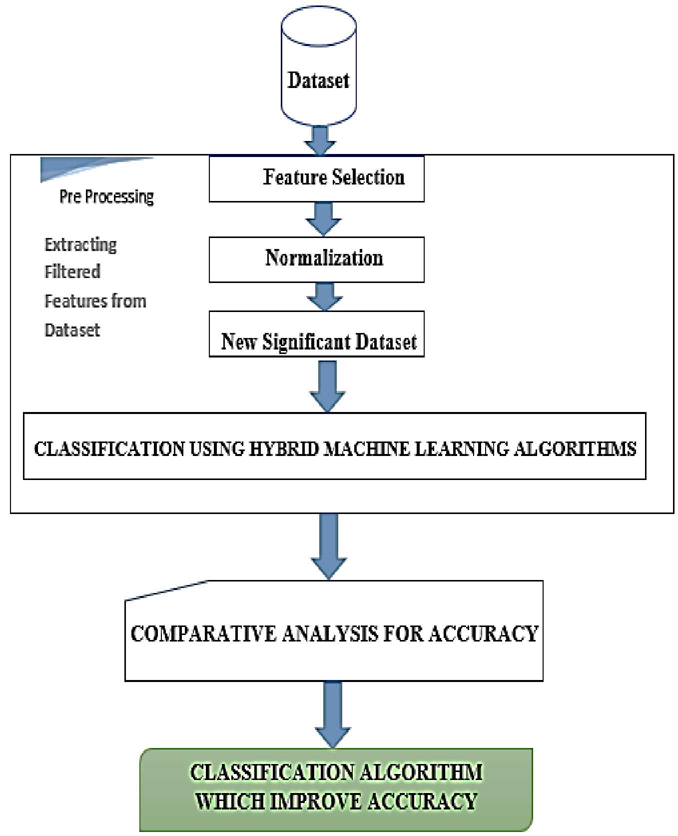


FIGURE 3 Hybrid Enhanced Machine learning process

The size of knowledge generated by various associations is increasing step by step exponentially. In web-based social networking, open and private evaluations of different topics or issues are communicated. Twitter is a small-scale blogging platform that enables individuals to express and articulate their points of view or send comments. Inquiry report on Facebook is a means of dissecting the customer's feeling by tweeting details (tweets). Analysis investigation refers to the use of natural language management, text analysis, and controlling electronic sensation investigation. Assumption Analysis also known as Opinion Mining (OM) aims at discovering the thoughts, mentalities and attitudes of individuals about a Stuff. The factor will talk to individuals, and multiple occasions. It gives us data on the positive, negative or unbiased extreme point and is the subject of an assumption.

A. STEPS FOR DATA LABELLING

In this phase we clean the data as well as label the data

- 0 for Poor
- 1 for Neutral
- 2 for Good

B. DATA PRE-PROCESSING TECHNIQUE

The following data pre pre-processing techniques are applied to alter raw reviews (data) to cleaned review (data), thus that it will become easier for us to ensure feature extraction in the next step.

- By removing html tags using Beautiful Soup
- by removing the non-character like digits or symbols
- By converting them to lower case

- By removing stop words such as "the" or "and" if needed
- By converting to roots of the words through stemming if needed

C. BAG OF WORDS (COUNTER)

The sentiment analysis of given text can be done in two ways. First, we need to find a word entrenching to convert a text into a numerical representation. Second, we fit the numerical representations of text to machine learning algorithms or deep learning architectures. In our case we only use machine learning, but deep learning can also be used in future work.

One common method to word transplanting is frequency-based enclosure such as Approach Bag of Words (Bow). Bow Approach can learn a comprehension list from a given amount and reflects each file on the basis of certain word counting methodologies. In this part, we will explore the Approach performance of using Bow with supervised learning algorithms. Workflow steps are mentioned in this part is shown in Fig 4.

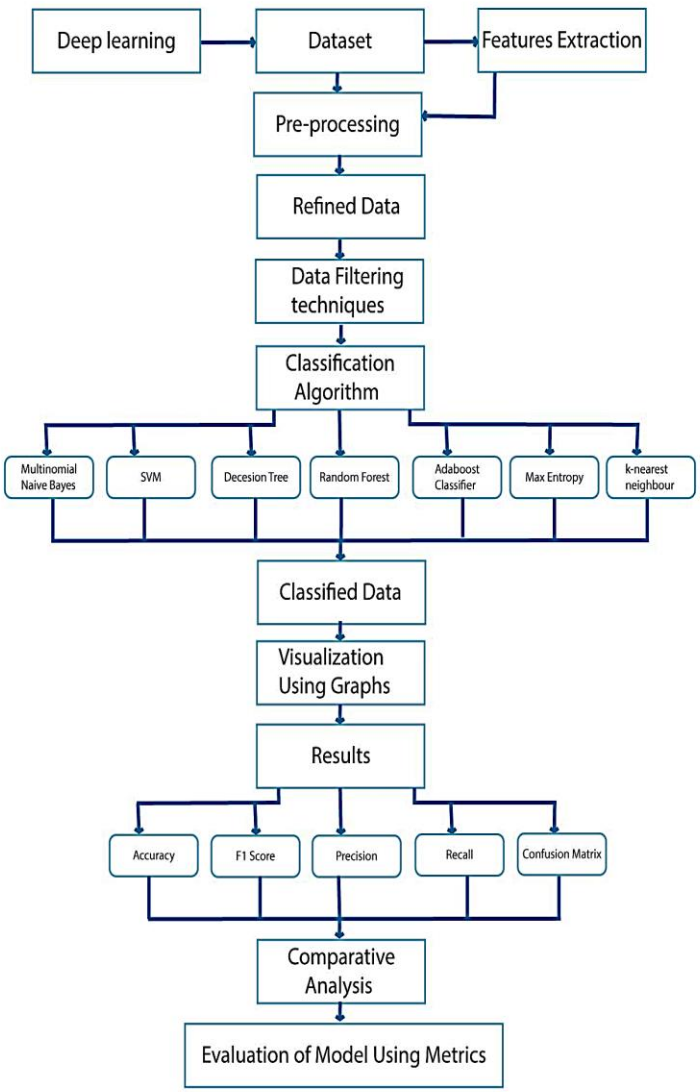


FIGURE 4 Proposed Hybrid Approach for Feature Extraction and Classification Using Machine Learning

Step 1: Data Pre-processing cleaned reviews (Imbalanced dataset) or raw reviews

Step 2: Create Bow (word bag) using the vectorizer Count / TF Idf in Sklearn.

Step 3: Convert text of the analysis into numerical representations (function vectors)

Step 4: Adapt vectors to supervised learning algorithms (e.g. Naive Bayes, Logistic Regression, etc.)

VI.CONCLUSION

There is a lot of working in machine learning now a days because it's emerging day by day and very demanding in terms to solve predictive analysis. Scenarios where prediction is required among different areas. Analysis is dependent on the group of individuals for the selection of algorithms or classifiers used for the novelty of dataset that is responsible for better results in less time. This is the most challenging task to do. Researchers have done a lot of work in this area, but still, there is a scope for machine learning techniques for the comparative aspects in the formation of hybrid approaching for the different classifiers showing the different result based on the datasets. Aim of this research is to provide a new technique for feature extraction and then selection based on data pre-processing techniques based on the deep learning classifier for best filtered data extractions for different selection of classifiers having the different results based on the preprocessing techniques and after applying specific algorithm as per the literature and knowledge to see the novelty based results.

Understandability is subjectively determined by the form of learning process and the scale of the resulting representation of the information. Hybrid methods increase the aspects ratio for improvements in predictive accuracy over standard understandable methods using generic Approach. Our hybrid Approach does classification using feature extraction in preprocessing of raw dataset by normalization for better classification, by removing unwanted noise from dataset either in the form of features or Data, to improve accuracy and minimization of Error Rate for the smaller amount of training of significant dataset and lesser amount of training time.

The hybrid comparative analysis needed for the selection of dataset with respect to machine learning classifier that will have different results with different data sets. Input with maximum accurate results is reproduced from our hybrid Approach shown in Tab. II. After that a comparative Analysis of different algorithms with different and same datasets can be made and that tell us about the accuracy and rate of misclassification Error shown in Tab. III. We Infer the results based on the different experiments conducted shown in Fig. 5, Fig. 6 and Fig. 7, if the raw data is being filtered and selective useable features are extracted so the accuracy will be increased based on the circumstances depending upon the novelty of dataset. A comparative Analysis shows the different results among the datasets for precision, accuracy, recall and f1 score against the generic techniques that are being modified by data preprocessing techniques for optimal results and also for the hybrid Approach for the best output result among the all.

V. COMPARATIVE ANALYSIS ON DIFFERENT DATASETS

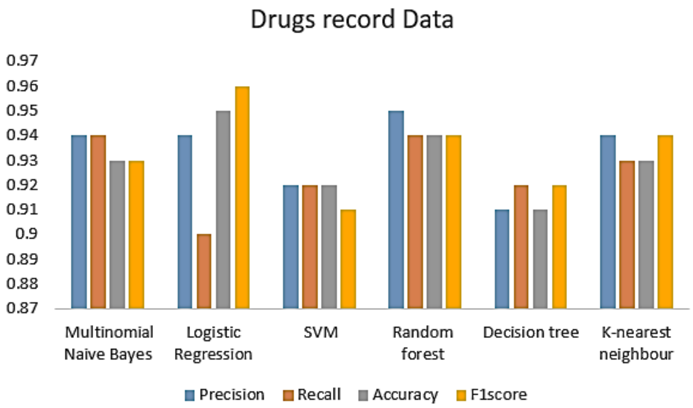
TABLE II: Comparative Analysis on different dataset with Hybrid approach

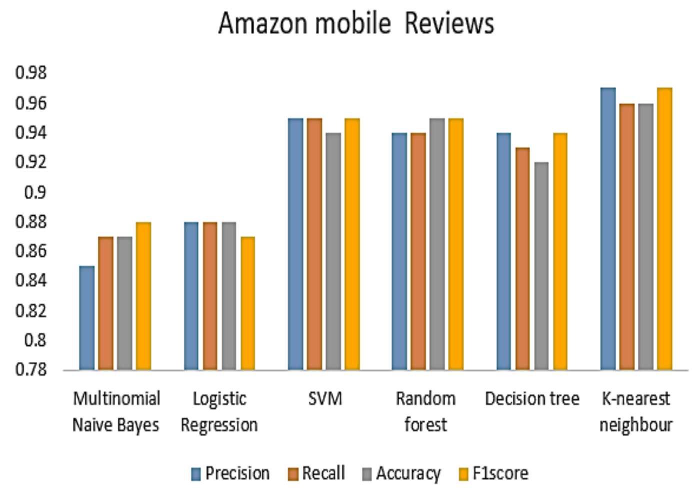
Sr No	About dataset	Approaches Used	Precision	Recall	Accuracy	F1score
1.	Amazon mobile Reviews	1. Multinomial Naive Bayes	0.85	0.87	0.87	0.88
		2. Logistic Regression	0.88	0.88	0.88	0.87
		3. SVM	0.95	0.95	0.94	0.95
		4. Random forest	0.94	0.94	0.95	0.95
		5. Decision tree	0.94	0.93	0.92	0.94
		6. K-nearest neighbor	**0.97**	**0.96**	**0.96**	**0.97**
2.	Breast Cancer Dataset	1. Multinomial Naive Bayes	**0.98**	**0.97**	**0.98**	0.97
		2. Logistic Regression	0.94	0.93	0.94	0.93
		3. SVM	0.92	0.92	0.93	0.92
		4. Random forest	**0.98**	**0.97**	0.97	**0.98**
		5. Decision tree	0.96	0.95	**0.98**	0.96
		6. K-nearest neighbor	0.94	0.92	0.95	0.94
3.	Amazon Alexa Products	1. Multinomial Naive Bayes	0.93	0.94	0.94	0.88
		2. Logistic Regression	0.93	0.93	0.94	0.89
		3. SVM	0.95	**0.95**	0.96	0.93
		4. Random forest	**0.96**	**0.95**	**0.97**	**0.95**
		5. Decision tree	0.93	0.94	0.94	**0.95**
		6. K-nearest neighbor	0.91	0.92	0.93	0.94
4.	Bilingual "Hindi-English" Data	1. Multinomial Naive Bayes	0.69	0.68	0.69	0.68
		2. Logistic Regression	0.64	0.63	0.64	0.62
		3. SVM	**0.70**	**0.70**	**0.70**	**0.70**
		4. Random forest	0.67	0.67	0.66	0.67
		5. Decision tree	0.69	0.68	0.69	0.68
		6. K-nearest neighbor	0.66	0.65	0.66	0.65
5	Drugs record Data	1. Multinomial Naive Bayes	0.94	0.94	0.93	0.93
		2. Logistic Regression	0.94	0.94	**0.95**	**0.96**
		3. SVM	0.92	0.92	0.92	0.91
		4. Random forest	**0.95**	**0.94**	0.94	0.94
		5. Decision tree	0.91	0.92	0.91	0.92
		6. K-nearest neighbor	0.94	0.93	0.93	0.94

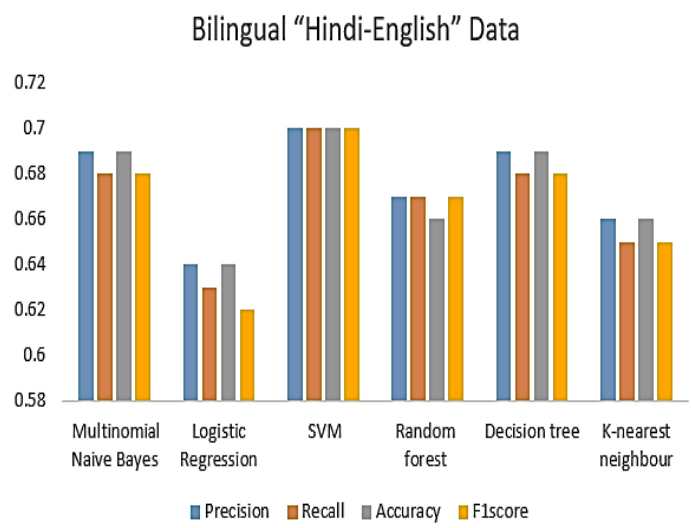
TABLE III: Comparative Analysis on different dataset with enhanced generic approach

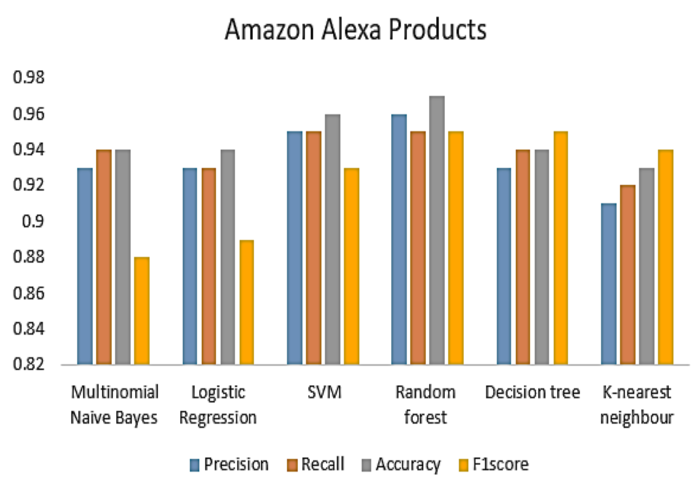
Sr No	About dataset	Approach's Used	Precision	Recall	Accuracy	F1score
1.	Amazon mobile Reviews	1. Multinomial Naive Bayes	0.84	0.86	0.86	0.85
		2. Logistic Regression	0.87	0.88	0.88	0.86
		3. SVM	0.94	0.93	0.94	0.93
		4. Random forest	0.93	0.94	0.93	0.93
		5. Decision tree	0.92	0.93	0.92	0.93
		6. K-nearest neighbor	**0.96**	**0.95**	**0.96**	**0.95**
2.	Breast Cancer Detection Data	1. Random forest	0.97	0.96	**0.97**	0.96
		2. Decision tree	0.93	0.92	0.94	0.94
		3. Naïve Bayes	0.91	0.92	0.93	0.92
		4. SVM	**0.98**	**0.97**	0.96	**0.97**
		5. Logistic Regression	0.96	0.94	0.97	0.95
		6. K-nearest neighbor	0.94	0.92	0.93	0.93
3.	Amazon Alexa Products	1. Naïve Bayes	0.93	0.92	0.92	0.88
		2. Logistic Regression	0.92	0.92	0.92	0.88
		3. Random forest	0.94	0.94	0.94	0.92
		4. SVM	**0.96**	**0.95**	**0.96**	**0.94**
		5. Decision Tree	0.93	0.92	0.93	0.92
		6. K-nearest neighbor	0.91	0.92	0.91	0.91
4.	Bilingual "Hindi-English" Data	1. Naïve Bayes	0.64	0.62	0.62	0.61
		2. K-nearest neighbor	0.64	0.62	0.62	0.61
		3. Random forest	0.64	0.62	0.62	0.61
		4. Logistic Regression	**0.64**	**0.62**	**0.62**	**0.61**
		5. Decision Tree	0.64	0.61	0.61	0.59
		6. SVM	0.63	0.62	0.61	0.60
5	Drugs Record Dataset	1. Random forest	0.92	0.91	0.92	0.92
		2. Decision tree	0.93	0.93	**0.94**	0.94
		3. Naïve Bayes	0.91	0.92	0.92	0.91
		4. SVM	**0.94**	**0.93**	0.93	**0.94**
		5. Logistic Regression	0.91	0.92	0.91	0.92
		6. K-nearest neighbor	0.92	0.92	0.93	0.93

VII.Experimental results









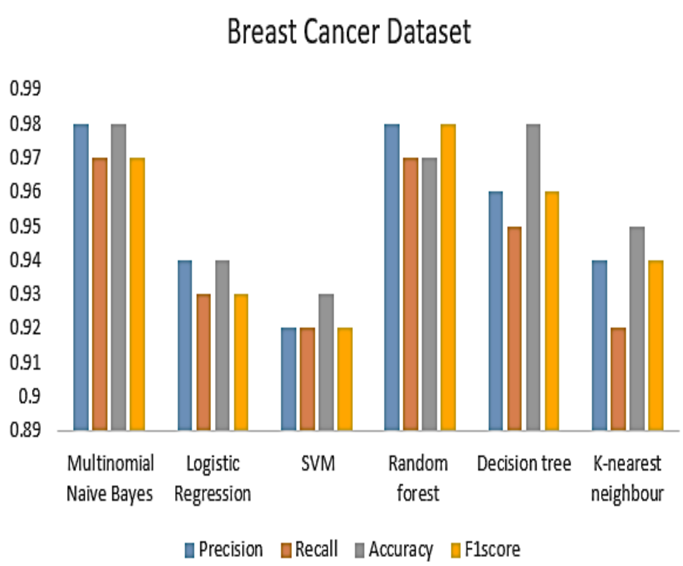


FIGURE 5: Experimental Results

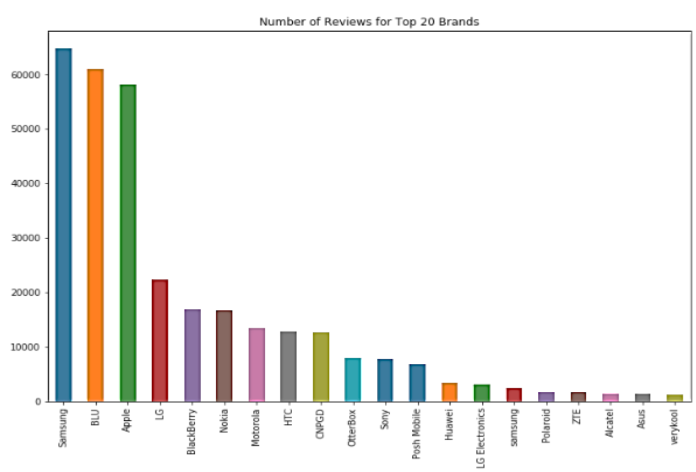


FIGURE 6: Number of Reviewers for top 20 brands

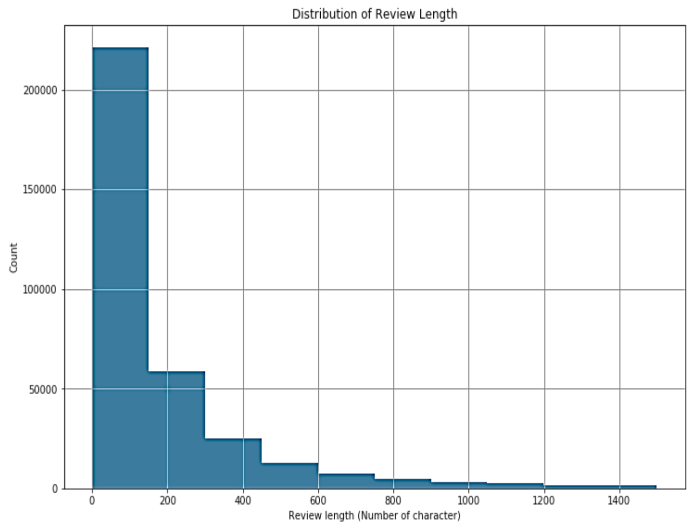


FIGURE 7: Experimental Results (reviewers length distribution)

REFERENCES

[1] C. Sammut & g. I. Webb, "encyclopedia of machine learning and data mining", springer, 2017

[2] Dalibor bužić *, jasminka dobša **, lyrics classification using naive bayes , mipro 2018.

[3] Ayman m mansou , texture classification using naïve bayes classifier, ijcsns international journal of computer science and network security, 2018

[4] Ianping gou and wenmo qiu, a local mean representation-based k-nearest neighbor classifier jjiangsu university, jiangsu key laboratory of security technology for industrial cyberspace, china. 2019

[5] Biau gerard and gerard biau. Journal of machine learning research 13 analysis of a random forests Approach, universite pierre et marie curie – paris 2012

[6] Amit gupte, sourabh joshi, pratik gadgul, akshay kadam, amit gupte et al , comparative study of classification algorithms used in sentiment analysis, / (ijcsit) international journal of computer science and information technologies, 2014

[7] Muhammad Nasir Khan, Syed K. Hasnain, Mohsin Jamil, Sameeh Ullah, "Electronic Signals and Systems Analysis, Design and Applications International Edition," in Electronic Signals and Systems Analysis, Design and Applications: International Edition , River Publishers, 2020

[8] Khan, Muhammad Nasir, Hasnain Kashif, and Abdul Rafay. "Performance and optimization of hybrid FSO/RF communication system in varying weather." Photonic Network Communications vol. 41, no. 1, pp. 47- 56, 2021.

[9] Khan, Muhammad Nasir, Mohsin Jamil, Syed Omer Gilani, Ishtiaq Ahmad, Muhammad Uzair, and H. Omer. "Photo detector-based indoor positioning systems variants: A new look." Computers & Electrical Engineering, vol. 83, pp. 106607, 2020.

[10] Kashif, Hasnain, Muhammad Nasir Khan, and Ali Altalbe. "Hybrid optical-radio transmission system link quality: link budget analysis." IEEE Access, vol. 8, pp. 65983-65992, 2020.

[11] Khan, Muhammad Nasir, and Fawad Naseer. "IoT based university garbage monitoring system for healthy environment for students." In 2020 IEEE 14th International Conference on Semantic Computing (ICSC), pp. 354-358. IEEE, 2020.

[12] Uzair, Muhammad, R. D. DONY, Mohsin Jamil, KHAWAJA BILAL AHMAD MAHMOOD, and Muhammad Nasir Khan. "A no-reference framework for evaluating video quality streamed through wireless network." Turkish Journal of Electrical Engineering & Computer Sciences, vol. 27, no. 5, pp. 3383-3399, 2019.

[13] Khan, Muhammad Nasir, Syed Omer Gilani, Mohsin Jamil, Abdul Rafay, Qasim Awais, Bilal A. Khawaja, Muhammad Uzair, and Abdul Waheed Malik. "Maximizing throughput of hybrid FSO-RF communication system: An algorithm." IEEE Access, vol. 6, pp. 30039-30048, 2018.

[14] Khan, Muhammad Nasir, Syed K. Hasnain, and Mohsin Jamil. Digital Signal Processing: A Breadth-first Approach. Stylus Publishing, LLC, 2016.

[15] Khan, Muhammad N. "Importance of noise models in FSO communications." EURASIP Journal on Wireless Communications and Networking vol. 2014, no. 1, pp. 1-10, 2014.

[16] Himani sharma,sunil kumar, a survey on decision tree algorithms of classification in data mining, april 2016,international journal of science and research (ijsr]

[17] Mohamed ahmed, ahmet rizaner, ali hakan ulusoy, a novel decision tree classification based on post-pruning with bayes minimum risk ahmed, yong deng, southwest university, china, 2018.

[18] Junhao wen, and cheng zhang, an improved random forest algorithm for predicting employee turnover xiang gao, hindawi mathematical problems in engineering , 2019.

[19] te han, dongxiang jiang comparison of random forest, artificial neural networks and support vector machine for intelligent diagnosis of rotating machinery, may transactions of the institute of measurement and control, 2018.

[20] Raphael couronné , philipp probst and anne-laure boulesteix , random forest versus logistic regression: a large-scale benchmark experiment , couronné et al. Bmc bioinformatics (2018).

[21] Nahid and Kong, "involvement of machine learning for breast cancer image classification: a survey," computational and mathematical methods in medicine, pp. 1-29, 2017.

[22] A jasper huang , fabio di troia , and mark stamp, coustic gait analysis using support vector machine, 2018.

[23] Li yueyang1, a and yi zhi wang1, b , detecting opinion polarities using ensemble of classification algorithms , iop conf. Series: journal of physics: 2019.

[24] Rukshan batuwita∗ and vasile palade† ∗ singapore-mit , class imbalance learning methods for support vector machines, alliance for research and technology centre;†university of oxford.2016

[25] Qingyao wu, mingkui tan, hengjie song, jian chen, and michael k. Ng, ml-forest: a multi-label tree ensemble method for multi-label classification, ieee transactions on knowledge and data engineering,2016

[26] Kim, K., et al. "General Labelled Data Generator Framework for Network Machine Learning." IEEE Xplore, 1 Feb. 2018

[27] Ahmad, Iftikhar, et al. "Performance Comparison of Support Vector Machine, Random Forest, and Extreme Learning Machine for Intrusion Detection." IEEE Access, vol. 6, 2018

[28] Naik, Amrita, and Lilavati Samant. "Correlation Review of Classification Algorithm Using Data Mining Tool: WEKA, Rapidminer, Tanagra, Orange and Knime." Procedia Computer Science, vol. 85, 2016

[29] Gislason, Pall Oskar, et al. "Random Forests for Land Cover Classification." Pattern Recognition Letters, vol. 27, no. 4, Mar. 2006

[30] Pondhu, Laxmi Narayana, and Govardhani Kummari. "Performance Analysis of Machine Learning Algorithms for Gender Classification." 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT), 2018

[31] Maxwell, Aaron E., et al. "Implementation of Machine-Learning Classification in Remote Sensing: An Applied Review." International Journal of Remote Sensing, vol. 39, no. 9, 2 Feb. 2018