JARTE
Journal of Applied Research in Technology & Engineering

4(2): 97-110, 2023
https://doi.org/10.4995/jarte.2023.19306
Received: February 24, 2023; Accepted: May 15, 2023; Published: July 31, 2023

PERFORMANCE EVOLUTION FOR SENTIMENT CLASSIFICATION USING MACHINE LEARNING ALGORITHM

Faisal Hassan [a], Naseem Afzal Qureshi [b], Muhammad Zohaib Khan [c], Muhammad Ali Khan [d], Abdul Salam Soomro [e], Aisha Imroz [f], Hussain Bux Marri [g]

[a] Department of Mathematics, Faculty of Science, University of Karachi, Karachi, Sindh, Pakistan.
[b] Department of Computer Science, Faculty of Science, University of Karachi, Karachi, Sindh, Pakistan.
[c] Software and Data Engineer, Shaheed Mohtarma Benazir Bhutto Institute of Trauma, Karachi, Sindh, Pakistan.
[d] Professor (Assistant), Industrial Engineering and Management, Mehran UET, Jamshoro, Sindh, Pakistan.
[e] Professor & Chairman, Industrial Engineering and Management, Mehran UET, Jamshoro, Sindh, Pakistan.
[f] Software Engineer, Avanza Solutions (Pvt.) Ltd, Karachi, Sindh, Pakistan.
[g] Professor (Meritorious) & Dean Faculty of Engineering Technology, BBSUTSD, Khairpur Mirs, Sindh, Pakistan.

[a] hassssanfaisal@gmail.com, [b] qureshinaseemafzal@gmail.com, [c] zohaibkhanpk2017@gmail.com, [d] muhammad.nagar@faculty.muet.edu.pk, [e] abdul.soomro@faculty.muet.edu.pk, [f] aishaimrozpk@gmail.com, [g] hussainbux@bbsutsd.edu.pk

Abstract:

Machine Learning (ML) is an Artificial Intelligence (AI) approach that allows systems to adapt to their environment based on past experiences. Machine Learning (ML) and Natural Language Processing (NLP) techniques are commonly used in sentiment analysis and Information Retrieval Techniques (IRT). This study supports the use of ML approaches, such as K-Means, to produce accurate outcomes in clustering and classification approaches. The main objective of this research is to explore the methods for sentiment classification and Information Retrieval Techniques (IRT). So, a combination of different machine learning algorithms is used with a dataset from amazon unlocked mobile reviews and telecom tweets to achieve better accuracy as it is crucial to consider the previous predictions related to sentiment classification and IRT. The datasets consist of user reviews ratings and algorithms utilized consist of K-Means Clustering algorithm, Logistic Regression (LR), Random Forest (RF), and Decision Tree (DT) algorithms. The amalgamation of each algorithm with the K-Means resulted in high levels of accuracy. Specifically, the K-Means combined with Logistic Regression (LR) yielded an accuracy rate of 99.98%. Similarly, the K-Means integrated with Random Forest (RF) resulted in an accuracy of 99.906%. Lastly, when the K-Means was merged with the Decision Tree (DT) Algorithm, the accuracy obtained was 99.83%.We exhibited that we could foresee efficient, effective, and accurate outcomes.

Keywords: machine learning; k-means; logistic regression; random forest; decision tree algorithms.

Cite as: Hassan, F., Qureshi, N.A., Khan, M.Z., Khan, M.A., Soomro, A.S., Imroz, A., Marri, H.B. (2023). Performance evolution for sentiment classification using machine learning algorithm. *J Appl Res Eng Technol & Engineering, 4*(2), 97-110. https://doi.org/10.4995/jarte.2023.19306

1. Introduction

An artificial model is fundamentally dependent on software programs that can acquire, understand, and decide concerning facts. Instead of simply obeying a collection of preset instructions, the program has the ability to find additional structures in data, identify relevant shapes, and understand methods and activities that enhance particular compensation. In other words, artificial systems acquire knowledge through training on vast amounts of information using various digital techniques. Artificial Intelligence (AI) and Machine Learning (ML) research also explore technologies that connect with human experts in evolving and altering physical and social situations. Primitive intelligent machines had minimal decision-making independence and assumed a stable work environment, performing the same interventions for the same problems. However, robotics has evolved to become an embedded device that can detect its surroundings and operate within the global environment to enhance specific objectives (Injadat et al., 2021; Reno, 2023; Riverside, 2023; Zhang et al., 2017). An intelligent model is everything that includes a functioning, sometimes not exclusively particular, pc with network access. Despite being sophisticated and proficient in complex data and information extraction, integrated systems often focus on activities related to the web server. However, the evolution of digital networks has influenced the utilization innovations and current offerings of prominent social collaboration platforms. In this context, the upcoming media plays an essential role in improving the gathering and interpreting of large volumes of data (Mata-Rivera et al., 2015; Sarker, 2021; Sheldon & Wigmore, 2023) (Gao et al., 2017; Shah et al., 2020). The advancement of digital approaches is unsurpassed, and user-generated content must become an integral part of daily life, particularly in institutions where most technology is changing the way learners participate and

*Corresponding author: M.Z. Khan, zohaibkhanpk2017@gmail.com

cooperate (Abad-Segura et al., 2020; Benavides et al., 2020; Tess, 2013). Communication networks' processors deliver detailed replies within a fixed period, and sales are not executed within the preceding text but rather shown by particular priority. It is important that these systems meet stringent quality standards, including usefulness, dependability, and accessibility (Golubic & Marusic, 1999; Matt et al., 2015; Vial, 2019).

Several ML systems for AI are in use today, especially in industries. Some require supervised learning, whereas others require unsupervised learning. The AI movement in this period has brought AI to global attention. A subcategory is ML. There are two types of circles: circular and square. We may feed the computer several kinds of data, and supervised learning works with structured or labelled data. Supervised learning applies what it has learned from previous data to new data (fresh data). We supply training data for the model, and the machine guesses whether it is a circle or a square. This supervised learning technique is utilized for subcategorization and regression. Unsupervised learning employs unstructured or disguised data and necessitates that the computer is fed a range of inputs. Unsupervised learning evaluates information before categorizing it; as a consequence, following classification, In each group, we collect data that is distinct from the others (Brownlee, 2016, 2019).

1.1. Comparison prediction method

We evaluate both supervised and unsupervised learning processes, as evaluated by K-Means, and the predictions approach created in LR, RF, and DT techniques, using the procedures of these approaches. These algorithms are used in clustering and classification strategies. These techniques alone supported a fast, accurate, and improved accuracy, precision, recall, f1-score, classifier, prediction, and clusters.

1.2. Scope

Nowadays, many researchers work in IoT sectors, with scientists seeking to better the innovative IoT system. This research enhanced and analysed several types of ML techniques by creating a classifier built on static analysis to identify sentiment exploration and IRT prediction. ML algorithms can be used in every IoT field, and still, ongoing research needs to enhance sentiment analysis and IRT. This study proposed an integrated approach for improved sentiment analysis using IRT prediction model accuracy, precision, recall, and f1-score used in different ML algorithms.

2. Background

2.1. Classification, regression, and clustering in machine learning

Classification is the systematic approach to grouping and subcategorization systems dependent on fundamental characteristics. In ML, there are various types of classifications. Classifiers LR, RF, and DT Classification systems involve prepared and classified data. Take, for illustration, Figure 1. It shows many classification elements used in various procedures.

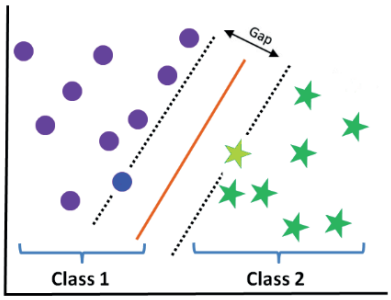

Figure1: Classification (Datavedas, 2018).

Although the association involving dependent and independent variables differs amongst regression models, linear and nonlinear regression techniques are focused on supervised and unsupervised learning strategies appropriately. It performs statistical problems.

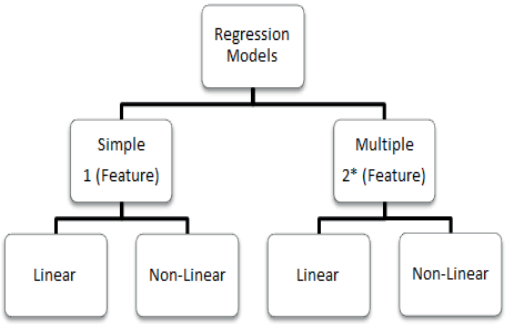

Figure 2: Regression Models.

ML approaches, both linear and nonlinear, use various predictive features from multiple sources, as illustrated in Figure 2. One common feature used in both linear and nonlinear analysis is regression analysis, which is the most frequent type of unsupervised method and has several applications in a variety of sectors. Another unsupervised method is clustering, which involves isolating and processing information to identify distinct groups or clusters within a dataset. By utilizing both regression analysis and clustering, ML approaches can effectively analyse complex data from multiple sources and provide valuable insights for a range of applications. The various clusters are shown in Figure 3.

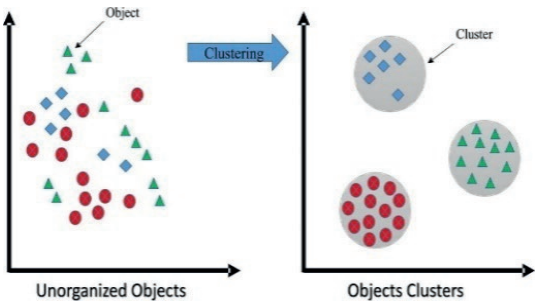

Figure3: Clustering (Abualigah et al., 2018).

2.2. Literature review

The following are the traditional approaches employed for sentiment classification. Jianqiang and Xiaolin (Jianqiang et al., 2018) used DCNN to generate word embeddings from large Twitter datasets using unsupervised learning. The polarity of sentiment score was used to combine these embeddings to create sentiment features for tweets (Alharbi & de Doncker, 2019; Jianqiang et al., 2018; Mansour, 2018). However, they did not extract the relationship between the semantics and improved feature information among words in the tweets. Ducange and colleagues (Ducange et al., 2019) developed a decision-making system to aid companies in managing their marketing campaigns. The system computes the reputation of brands and provides feedback by analysing the digital marketing campaigns and reviews. Sentiments of users are estimated based on positive, negative, or neutral polarities (Bouazizi & Ohtsuki, 2018; Ducange et al., 2019; Munjal et al., 2018). The approach did not succeed in incorporating a sentiment analysis aspect. Ohtsuki (Bouazizi & Ohtsuki, 2018) developed a Sentiment Quantification method to analyse sentiments and opinions from social media posts. The method used multi-class sentiment analysis to detect the sentiment polarity of users. The method used conventional multi-class classification on Twitter datasets and assigned scores to each sentiment (Bouazizi & Ohtsuki, 2018; Ducange et al., 2019). However, it did not consider sentiments across different polarities. In their study, Gao et al. created an IAAT model to categorize sentiments, which utilized an inverted index to be compatible with search engines already in use (Gao et al., 2017). The method's speed was attributed to the quick insertion and appending of operations. However, it proved to be inadequate when confronted with diverse retrieval constraints (Gao et al., 2017; Li et al., 2017; Mataoui et al., 2015; Schütze et al., 2008; Virmani et al., 2018). Bouazizi and Ohtsuki devised a Multi-class sentiment technique that employed both ternary and binary classifications. The approach was flexible and could be used to categorize texts into various classes (Bouazizi & Ohtsuki, 2017; Hassan et al., 2017; Jianqiang et al., 2018). The method selected features to execute sentiment classification. Spider monkey optimization algorithm (Bansal et al., 2014) is also a swarm-based algorithm benefiting from the division and composition of spider monkeys in nature. Kumar et al. introduced non-linear perturbation rate in Spider Monkey Optimization (SMO) based on exponential function (Kumar, Nayyar, et al., 2020), bag-of-features (Kumar et al., 2021) and hyperopic function (Kumar, Sharma, et al., 2020). Sharma et al. (2016) introduced the concept of aging in SMO and considered individuals' age. Munjal et al. proposed a model for opinion dynamics (Munjal et al., 2017, 2019). Many researchers used the Spider Monkey Optimization (SMO) technique for optimiation and analysis (Bouazizi & Ohtsuki, 2017, 2018; Iqbal et al., 2019; Singhal, 2001).

Baig et al. investigated the efficacy of machine learning (ML) techniques for predicting the academic success of Higher Education Institutions' (HEIs) students. Machine Learning (ML), deep learning, Fuzzy C-Means, Multi-Layer Perceptron (MPL), Logistic Regression (LR) and Random Forest (RF) algorithms were investigated separately and in combinations for predicting student achievement in the classroom. These strategies were also assessed using quantitative standards including accuracy, detection rate, and false alarm rate (Baig et al., 2023). Channar et al. highlighted knowledge sharing factors in the Higher Education Institutions (HEIs) (Channar et al., 2023). Information seeking behaviour is changing dynamically (Rahoo et al., 2019, 2020) and the information retrieval patterns of the students and faculty members in Higher Education Institutions (HEIs) are also change. Modern technologies like Artificial Intelligence (AI) & Machine learning (ML) promoted the effective use of Social Media (Buriro et al., 2018) also in Information seeking/retrieval in Higher Education Institutions (HEIs) (Memon et al., 2020; Nagar, Kalhoro, et al., 2018; Nagar, Rahoo, et al., 2018). Khan et al. conducted the performance analysis of machine learning (ML) and data mining techniques for anomaly detection in credit card frauds. Principal Component Analysis (PCA), data mining, Fuzzy C-Means methodologies, Logistic Regression (LR), Decision Tree (DT) and Naive Bayes (NB) algorithms have been reviewed and compared (M.Z. Khan et al., 2023).

Zaman et al. compared the comparative study of Relational Database Management Systems (RDBMS) and non-relational database management systems i.e. Azure SQL and Atlas Mongodb NoSQL databases. Loading time, response time, and retrieval time of both Azure SQL and Atlas Mongodb NoSQL databases were compared their speed, efficiency and performance were monitored (Zaman et al., 2021). Khan et al. compared the comparative study of Relational Database Management Systems (RDBMS) and non-relational database management systems i.e. SQL and NoSQL databases. Both were compared in terms of loading, response, and retrieval times to discover their smoothness, efficiency and performance (M.Z. Khan et al., 2022). Khan et al. applied the Artificial Intelligence (AI), Machine learning (ML) and Principal Component Analysis (PCA) to evaluated the performance computation between K-Means, K-nearest neighbor (KNN), support vector machine (SVM) algorithms. Authors have also presented the simulations that show the estimated criteria, parameters and efficiency with effective algorithms' performance and state-of-the-art results (Khan et al., 2022). Many organizations in Pakistan lags behind in the optimization of operations (Arain et al., 2020; Kalwar & Khan, 2020a)cost and process efficiency. Stitching in the footwear is considered as the bottleneck of the production because of the variability of operations in the different articles (shoes by applications of Artificial Intelligence (AI) & Machine learning (ML). The recent applied case studies of Pakistani organizations in the context of optimization by automation include procurement report (Kalwar & Khan, 2020), routine report making (Kalwar et al., 2020), purchase order (Kalwar & Khan, 2020b), acquisition report (Kalwar & Khan, 2020), planning report (Kalwar et al., 2021), Supplier Price Evaluation Report (M.A. Khan, Kalwar, Malik, et al., 2021), material delivery time analysis (M.A. Khan, Kalwar, & Chaudhry, 2021) product mix & profit maximization (Kalwar, Khan, et al., 2022), and order costing analysis (Kalwar, Shahzad, et al., 2022). The demand management system at logistic centre of civil aviation authority (CAA) Karachi by using VBA which reduced 50% of working time (Chaudhry et al., 2021). Kalwar & Khan saved Seventy-five percent (75%) of the staff time which would have otherwise spent manually

creating the procurement report was saved by automation in operations. The process of generating the order costing report was significantly faster after automation compared to the manual technique, with a time reduction of 85.92% (Kalwar & Khan, 2020b). More recently, the automation was implanted at material cost comparative analysis (MCCA) in an industry which resulted in 100% accuracy and 72.20% time reduction in 58.51 minutes (Kalwar et al., 2023).

2.3. Performance analysis

Python is a high-level scripting language mainly used for programming and machine learning techniques, website development, and databases. We utilized the Anaconda Navigator ->Jupyter Notebook GUI framework. We used Python to link datasets and perform K-Means, LR, RF, and DT algorithms. The dataset is for amazon-unlocked-mobile-reviews and telecom-tweets and prediction based on user reviews rating. There are seven attributes (columns) and 413840 tuples (rows) in this dataset. We ran three separate programs, the first of which used K-Means, and Logistic Regression (LR) techniques, the second of which used the K-Means, and Random Forest (RF) method, and the third of which used K-Means, and Decision Tree (DT) procedure, and all of which utilized the same dataset. On the personal computer, all of these apps are running. The following is the computer's configuration:

- Second Generation Intel (R) Core (TM) i5-2520M CPU @ 2.50 GHz.
- RAM of 4.00 GB.
- The system is a 64-bit operating system.
- Windows 10 (Home).
- 500 GB hard disk.

3. Data collection

Collected data from the Kaggle website, which covers a large number of machine learning datasets and the dataset amazon-unlocked-mobile-reviews and telecom-tweets. Previously, which has been used by authors (Chugh et al., 2021). This dataset includes 413840 tuples and seven attributes—every attribute on behalf of intelligent mobile specifications, tweets, and reviews measures. With one target class which signifies the status of each outcome separately and in this dataset, a total of 290744 negative reviews results and 123101 positive reviews results (Kaggle, 2023). Table 1 briefly describes the characteristics and variables included in the implementation dataset for this study, which were Amazon-unlocked-Mobile-Reviews and Telecom-Tweets-Dataset.

It also aims to explore data normalization, pre-processing, simulation and induction requirements, critical criteria, complexity issues, identified create post-processing, simulations, system effectiveness, and the most significant phases through processes described. To begin with, facts are retrieved from the dataset. After that data is prepared, the data is pre-processed and standardized in the last stage. Table 2 displays the cleaned and pre-processed dataset. Figure 4 depicts the complex information graphically without K-Means execution. Purple is the X value color circle, while yellow is the second Y value color circle.

Table 1: Actual Amazon-unlocked-Mobile-Reviews and Telecom-Tweets-Dataset.

Product Name	Brand Name	Price	Rating
"CLEAR CLEAN ESN" Sprint EPIC 4G Galaxy SPH-D700*FRONT CAMERA*ANDROID*SLIDER*Q WERTY KEYBOARD*TOUCH SCREEN	Samsung	199.99	5
"CLEAR CLEAN ESN" Sprint EPIC 4G Galaxy SPH-D700*FRONT CAMERA*ANDROID*SLIDER*Q WERTY KEYBOARD*TOUCH SCREEN	Samsung	199.99	4
"CLEAR CLEAN ESN" Sprint EPIC 4G Galaxy SPH-D700*FRONT CAMERA*ANDROID*SLIDER*Q WERTY KEYBOARD*TOUCH SCREEN	Samsung	199.99	5
"CLEAR CLEAN ESN" Sprint EPIC 4G Galaxy SPH-D700*FRONT CAMERA*ANDROID*SLIDER*Q WERTY KEYBOARD*TOUCH SCREEN	Samsung	199.99	4
Reviews	**Review Votes**	**Class**	
I feel so LUCKY to have found this used (phone to us & not used hard at all), phone on line from someone who upgraded and sold this one. My Son liked his old one that finally fell apart after 2.5+ years and didn't want an upgrade!! Thank you Seller, we really appreciate it & your honesty re: said used phone.I recommend this seller very highly & would but from them again!!	1	Positive	
nice phone, nice up grade from my pantach revue. Very clean set up and easy set up. never had an android phone but they are fantastic to say the least. perfect size for surfing and social media. great phone samsung	0	Negative	
Very pleased	0	Negative	
It works good but it goes slow sometimes but its a very good phone I love it	0	Negative	

Table 2: Pre-processed Amazon-unlocked-Mobile-Reviews and Telecom-Tweets-Dataset.

Product Name	Brand Name	Price	Rating
0.151451	0.913760	0.953259	0.715392
0.743204	0.969356	0.461230	0.621011
0.216024	0.850044	0.313211	0.521588
0.266557	0.033443	0.383097	0.899824
Reviews	Review Votes	Class	
0.446439	0.672079	0.644045	
0.313483	0.608386	0.691252	
0.556572	0.325855	0.288372	
0.204689	0.089192	0.274413	

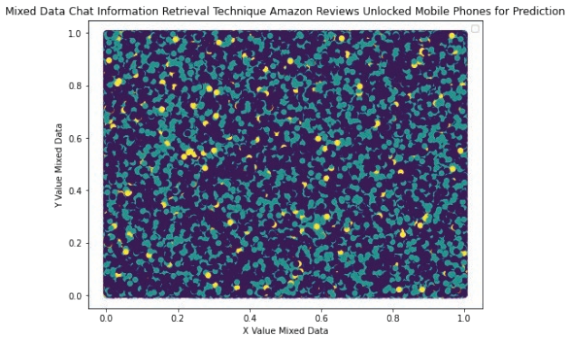

Figure 4: K-Means Earlier Execution Mixed Data Chart.

3.1. K-Means Clustering Algorithm

Clustering is the most popular kind of unlabelled data, having numerous applications and implementations in several industries. It's The act of separating and handling data on behalf of a pc, and It's the collection of information that outcomes from this technique. Every cluster is allotted unique identifier (ID) values. The unsupervised K-means technique is an ML method. This approach classifies data as either mixed or unstructured. The dataset initiates with the first group with specific randomly average scores, which serves as the starting point for each subsequent group. It then repeats computations to enhance the location of the intermediate values (Dabbura, 2018). The K-means algorithm operates on the following fundamental principles:

1. Measure the number of K clusters.
2. Calculate the centroids by sorting the collection and then randomly choosing the centroids' K values.
3. Identifying cracks once there is no development within the centroids. In the other sense, the method of clustering data continues unaltered.
4. Determine the number of patterned lengths between the information's focal facts and all centroids.
5. Identify each data item by indicating the closest correct number (centroid significance).
6. Calculate the sum of all data emphases allocated towards each group to get the cluster centroids.
7. End.

Those were some scientific splitting methods and expressions applied to every program. This approach is based on categorization, like the Euclidean, Manhattan, and Hamming measures. Some of these approaches are as follows:

Euclidean $$\sqrt{\sum_{i=1}^{k}(xi-yi)2}$$

Manhattan $$\sum_{i=1}^{k}|xi-yi|$$

Minkowski $$\left(\sum_{i=1}^{k}(|xi-yi|)q\right)1/q$$

We incorporated a standardized collection in this pre-processing and used it to create mixed data representations. K-Means is a method for filtering and preparing large datasets to improve comprehensibility while eliminating data redundancy. K-Means was used to identify two clusters. K-Means clustering is a fast approach in which each piece of information is assigned a mathematical probability or possibility score indicating whether it corresponds to that cluster. It accomplishes this by successively generating new statistically independent values that enhance variance. This logic was instantly applied in the proposed matrix to form a membership matrix displaying the degree of connection between the sample and each cluster. This approach employs a clustering technique, as evidenced by the centroid clustering values, the K-Means algorithm, and a ten-dimensional dataset. The amazon-unlocked-mobile-reviews and telecom-tweets dataset include one column cluster. It's explicitly defined as an eight-dimensional dataset, with two features based on user reviews rating, dataset pre-processing values, and one feature target property cluster number (see Table 3). The K-Means Clustering Centroid Value briefly defined is given below. Figure 5 shows the two clusters after the unstructured dataset was converted to structured data. These graphs depict purple and teal sets with centroid values specified by the yellow star (*). Figure 6: K-Means determined sum of squared error line chat.

K-Means Clustering Centroid Value.

Array([[0.4989848, 0.49929046, 0.49886558, 0.74988587, 0.50280366, 0.4985661, 0.50177629],

[0.49985799, 0.50064174, 0.50200912, 0.2487589, 0.49739375, 0.50063244, 0.49831234]])

Table 3: K-Means Two Clusters Pre-processed Amazon-unlocked-Mobile-Reviews and Telecom-Tweets-Dataset.

Product Name	Brand Name	Price	Rating
0.151451	0.913760	0.953259	0.715392
0.743204	0.969356	0.461230	0.621011
0.216024	0.850044	0.313211	0.521588
0.266557	0.033443	0.383097	0.899824
0.847414	0.361462	0.381868	0.142174
Reviews	Review Votes	Class	Clusters
0.446439	0.672079	0.644045	0
0.313483	0.608386	0.691252	0
0.556572	0.325855	0.288372	0
0.204689	0.089192	0.274413	0
0.491633	0.074825	0.845570	1

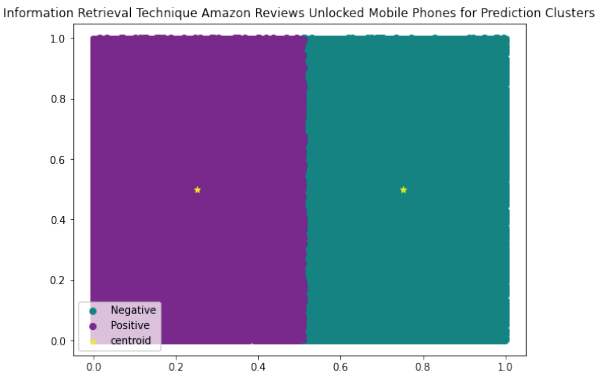

Figure 5: K-Means Two Clusters.

Advancement Of CFD exhausting equally clustering techniques is associated with using precision, recall, and f-measure.

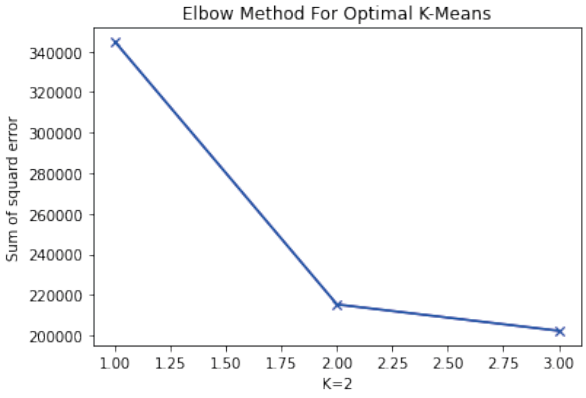

Figure 6: K-Means Sum of Squared Error Line Chart.

A suspicion score is calculated according to the extent of deviation from the standard patterns. Therefore, based on user reviews, the rating prediction result is classified as valid or suspect or predictions.

3.2. Logistic Regression (LR) Algorithm

The LR is a classification application that utilizes ML. It is indeed a probabilistic hypothesis-based forecast analysis method. A Linear Regression is similar to the LR model. Nevertheless, LR applies a much more complicated differential equation, also defined as the 'Sigmoid function' or the 'regression model,' instead of a linear transformation. The LR hypothesis limits the mathematical problem to integer coefficients 0 and 1. As a result, functions fail to explain it until this could have a meaningful value than one or even less than zero, which is not possible under the logistic regression assumption. Subsequently, linear methods attempt to represent it. Because this could have a value greater than one, it will be less than zero, which is impossible under the LR hypothesis (Pant, 2019). Those were some specific measures or procedures which can be used in any method. This procedure is based on classifications, such as the Sigmoid, Linear, Cost Linear, and Nonlinear Logistic Regression. A few of these techniques are as follows:

Sigmoid	$\mathbf{S}(\boldsymbol{z}) = \frac{\mathbf{1}}{(\mathbf{1} + \mathbf{e}^{-\mathbf{z}})}$
Linear Regression	$\boldsymbol{y} = \boldsymbol{e}\text{^}(\boldsymbol{b0} + \boldsymbol{b1} * \boldsymbol{x}) / (\mathbf{1} + \boldsymbol{e}\text{^}(\boldsymbol{b0} + \boldsymbol{b1} * \boldsymbol{x}))$
Cost Linear Regression	$(\boldsymbol{Cost}\,(\boldsymbol{h\theta}(\boldsymbol{x}), \boldsymbol{y})) = -\boldsymbol{log}\,(\boldsymbol{h\theta}(\boldsymbol{x})), \boldsymbol{if}\ \boldsymbol{y} = \mathbf{1}\ \boldsymbol{and}$ $(\boldsymbol{Cost}\,(\boldsymbol{h\theta}(\boldsymbol{x}), \boldsymbol{y})) = -\boldsymbol{log}\,(\mathbf{1} - \boldsymbol{h\theta}(\boldsymbol{x})), \boldsymbol{if}\ \boldsymbol{y} = \mathbf{0}$
Nonlinear Regression	$\mathbf{Y} = \mathbf{f}(\mathbf{X}, \boldsymbol{\beta}) + \boldsymbol{\varepsilon}$

The basic steps of the LR algorithm are as follows:

1. Initialize all parameters *(B_0, B_1, etc.)*.
2. Compute (predict) dependent variable *($h_\theta(x)$)*.
3. Compute cost function *(Cost($h_\theta(x)$,y)) or* an Logistic Regression function.
4. Compute gradient for the cost function.
5. Update all parameters.
6. Repeat steps 2 to 5.
7. End.

The LR prediction labels the old data values and predicts the value of the new data, where we try to make your predictions fit the labels during preparation; Figure 7 shows the result of that LR confusion matrix.

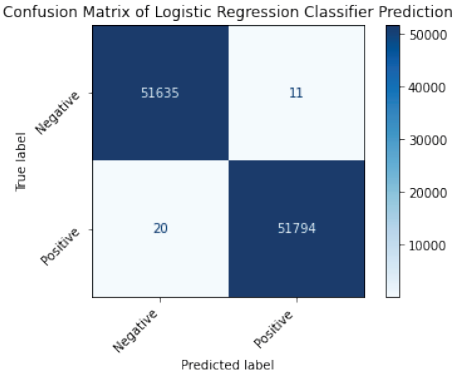

Figure 7: Confusion Matrix Logistic Regression (LR)Algorithm.

As of this paper, the confusion matrix has the meaning as [[A,B][C,D]].

- A is the digit of accurate expectations that illustration is negative,
- B is the number of erroneous forecasts that instance is positive,
- C is the digit of negative indecentextrapolations, and,
- D is the number of exact optimisticestimates.

On a synthetic dataset, we can show this by plotting the ROC curve for a no-skill classifier and LR algorithm Receiver Operating Characteristic (ROC Curve); Figure 8 shows the result.

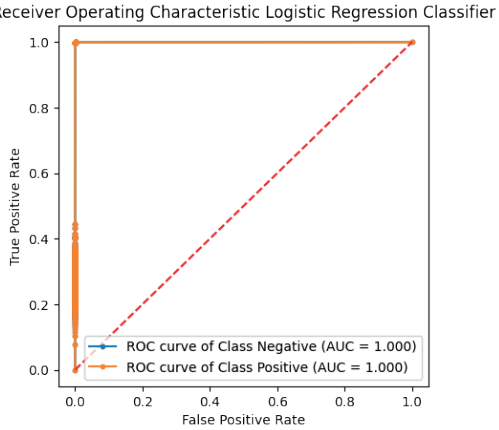

Figure 8: Logistic Regression (LR) Receiver Operating Characteristic (ROC Curve).

In this paper, we used the concept of the ROC curve to analyse our model. This will gain the accuracy of the estimated prediction method while frequently meeting user reviews rating prediction patterns.

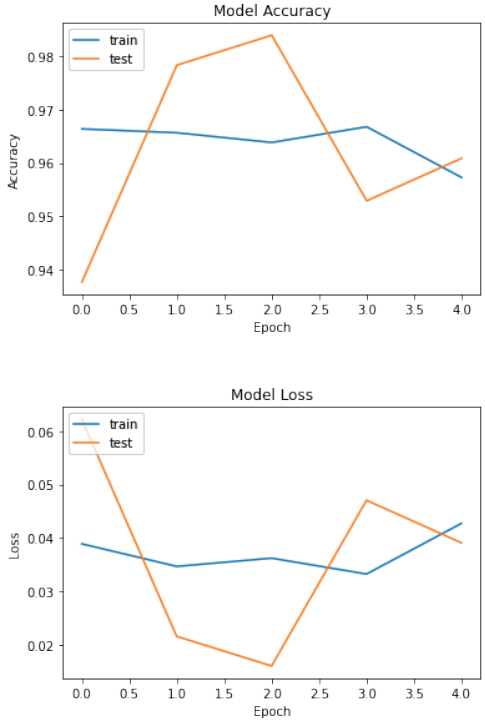

Figure 9, 10: Model Accuracy and Model Loss Logistic Regression (LR) Algorithm.

This studyevaluated our model using the model accuracy and model loss concepts. This will improve the correctness of the approximated prediction approach while also regularly fulfilling user reviews rating prediction patterns: Figures 9 and 10 display the resulting model accuracy and model loss Logistic Regression (LR) algorithm.

3.3. Random Forest (RF) Algorithm

The RF is an ML technique that is used to resolve classifier issues. It uses different classifiers, a sort of difficult resolving system that uses classifying approaches. That's the merging of numerous categories to solve complicated problems and enhance the system's effectiveness. The RF approach, built on classification tree predictions, decides the efficacy. It makes assumptions by approximating or calculating the results of numerous trees. The quality of the output increases even as the number of nodes increases. The limitations of a DT are reduced by exhausting RF (Boateng et al., 2020; Mbaabu, 2020; Tutorialspoint, 2023).

The RF process is based on the following fundamental principles:

1. Initiate by randomly picking observations through facts.
2. The program will instead create a tree structure for every instance. The outcomes for every tree structure will then be produced.
3. Every generated that came as a result would be selected or decided on throughout this stage.
4. Lastly, select the most preferred forecasting outcome as unique of the most predicted results.

Moreover, the RF Algorithm approach utilizes specific mathematical functions or formulas. This approach, as well as Mean Squared Error (MSE), Gini (Coefficient, Index, or Ratio), and Entropy (Nitze et al., 2012; Reis et al., 2018; Schott, 2019) may be used by the RF Algorithm. Evaluate the following procedures as examples:

Mean Squared Error (MSE)

$$\frac{1}{N} = \sqrt{\sum_{i=1}^{N} (xi - yi)^2}$$

Gini Coefficient

$$Gini = 1 \sum_{i=1}^{C} (p_i)^2$$

Entropy

$$Entropy = \sum_{i=1}^{c} -p_i * log_2(p_i)$$

With the use of the RF algorithm, The LR prediction labels the old data values. It predicts the importance of data, where we try to make your predictions fit the labels during preparation using the algorithm RF algorithm Figure 11 shows the result of that confusion matrix.

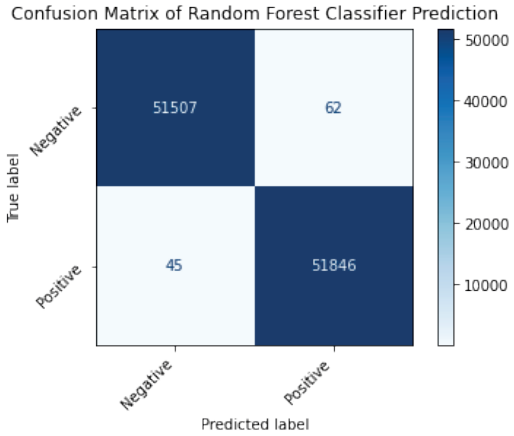

Figure 11: Confusion Matrix Random Forest (RF) Algorithm.

Assuming that our RF model sounds fitted in this, the confusion matrix helped us calculate the predicted labels of our prediction.

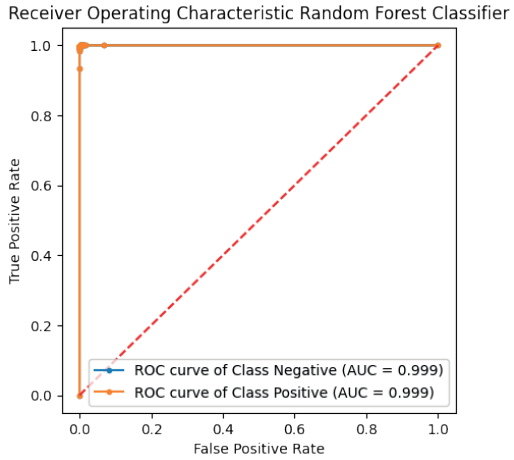

Figure 12: Random Forest (RF)Receiver Operating Characteristic (ROC Curve).

It displays the systematic performance of a classifier model when its discriminating inception is changed. ROC analysis is intrinsically connected to cost/benefit research in rational decision-making, and the RFROC Curve in Figure 12 illustrates the result.

This work evaluated our model using the model accuracy and model loss concepts. This will improve the approximated prediction strategy's accuracy while also regularly fulfilling user reviews rating forecast patterns. Figures 13 and 14 present the result of the model accuracy and model loss RF algorithm.

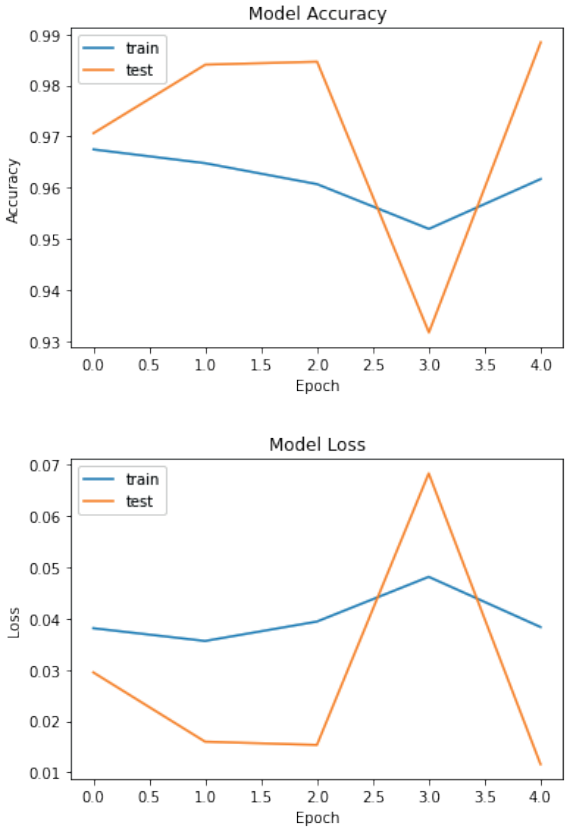

Figure 13, 14: Model Accuracy and Model Loss Random Forest (RF)Algorithm.

3.4. Decision Tree (DT) Algorithm

The DT method is a type of supervised machine learning method. Furthermore, the decision tree approach can be used to outperform the classification technique's functionalities. The goal of building a Tree Structure is to create an evolution model for predicting the categories or variables of the target variable in terms of the primary selection rules learned from previous outcomes (training data). To construct a data point for a fact-based decision Tree, we start at the tree's base. The significance of the integrity of the data is evaluated against the quantity of the essential attribute. We proceed toward the next node by analysing the branches corresponding to that discovery. Some analytical procedures or expressions can be combined with any program (Chauhan, 2020). This approach, like the Information Gain and Entropy methods, is established on creating a DT and classification. The following are some of these approaches:

Information Gain

$$\textbf{Information Gain} = \mathbf{E(Y)} - \mathbf{E(Y|X)}$$

Entropy

$$\boldsymbol{Entropy(s)} = -\boldsymbol{P}(+)\boldsymbol{log2}\ \boldsymbol{P}(+) - \boldsymbol{P}(-)\boldsymbol{log2}\ \boldsymbol{P}(-)\textbf{Information}$$

The DT algorithm's significant steps work as follows:

1. It started with S, the initial root node.
2. The procedure repeats over the comparatively same variable of the sequence, evaluating the Entropy (H) and Information Gain (IG) of this capability with each iterative process.
3. Instead, the feature with the shortest entropy or most significant information gain is selected.
4. The measurement S is instead partitioned by such a desired characteristic to provide many observations.
5. The approach is repeated for each segment, reflecting only qualities that have never been accepted previously.
6. End.

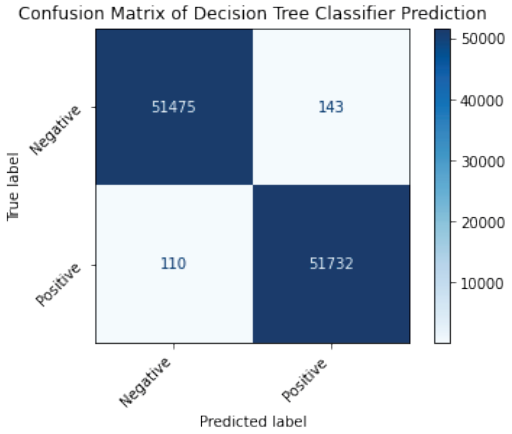

Figure 15: Confusion Matrix Decision Tree (DT) Algorithm.

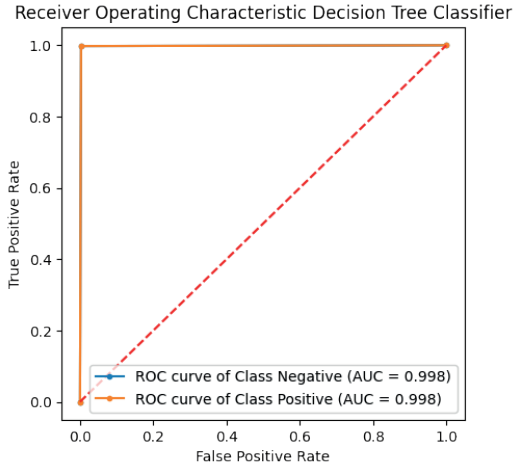

Figure 16: Decision Tree Receiver Operating Characteristic (ROC Curve).

With the use of the DT algorithm, The LR prediction labels the old data values. It predicts the importance of data, where we try to make your predictions fit the labels during preparation using the algorithm decision tree Figure 15 shows the result of that confusion matrix.

Assuming that our decision tree model sounds fitted in this, the confusion matrix helped us calculate the predicted labels of our detection prediction.

It displays the systematic performance of a classifier model when its perceptive beginning is changed. ROC analysis is intrinsically connected to cost/benefit research in rational decision-making, and the DT ROC Curve Figure 16 illustrates the result.

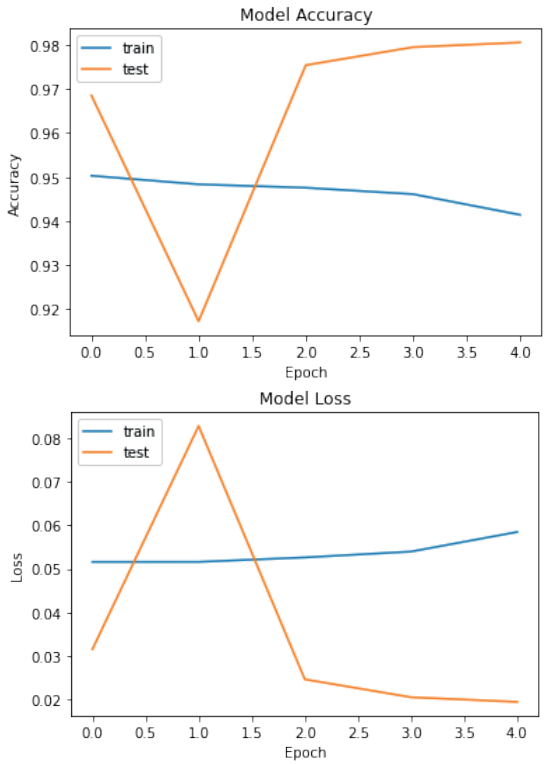

Figure 17, 18: Model Accuracy and Model Loss Decision Tree Algorithm.

This work estimated our model using the model accuracy and model loss concepts. It will improve the approximated prediction strategy's accuracy while regularly fulfilling user reviews rating forecast patterns —Figures 17 and 18 present the resulting model accuracy and model loss decision tree algorithm.

4. Results and Discussion

Machine learning is a practical technique that teaches algorithms to solve challenges instead of supervised by ML. DL is currently the ML champion, thanks to better processes, computer power, and large data sets. Nevertheless, conventional ML techniques continue to play an essential role in the industry. This paper used a new approach based on the integration K-Means. Then, a comparative study of the performance was performed, including other supervised machine learning algorithms, to reach the best classifier. Logistic Regression (LR), Random Forest (RF), and Decision Tree (DT) were the supervised machine learning algorithms used in this study. In the next step, we combine the algorithms to check the most accurate combination of algorithms prediction based on user reviews rating. The highest accuracy we gained using our algorithms' combination was through K-Means,

Logistic Regression (LR). Next, we acquired the second-ranked combination, K-Means, Random Forest (RF). Finally, K-Means, Decision Tree (DT) was ranked third.

Table 4: Combination of Algorithms Model Accuracy Based on User Reviews Rating.

Combination of Algorithms	Accuracy
K-Means,Logistic Regression (LR).	99.98066885752948 %
K-Means,Random Forest (RF).	99.90624395901798 %
K-Means,Decision Tree (DT).	99.8366518461241 %

Table 5: Combination of Algorithms Parameter Score Prediction Based on User Reviews Rating.

S/#	Parameter Score (%)	K-Means, Logistic Regression (LR)	K-Means, Random Forest (RF)	K-Means, Decision Tree (DT)
1	Accuracy	0.99980668	0.99906243	0.99836651
2	Precision	0.99980668	0.99906243	0.99836651
3	Recall	0.99980668	0.99906243	0.99836651
5	Sensitivity	0.99980678	0.99913006	0.99841165
6	Specificity	0.99980659	0.99899481	0.99832156
7	F1-Score	0.99980668	0.99906243	0.99836651

The combination of our executed algorithm, K-Means, and Logistic Regression (LR), has reached its maximum outcome in terms of accuracy. However, the cross between K-Mean, and Random Forest (RF) was ranked second, followed by our third algorithm consisting of K-Means, Decision Tree (DT) Algorithm, and Tables 4 and 5 briefly defined results.

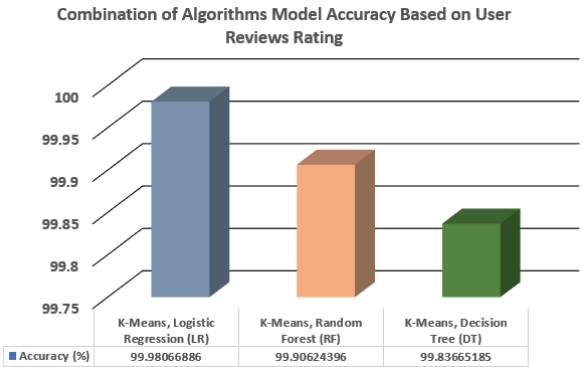

Figure 19: Combination of Algorithms Graph Model Accuracy Based on User Reviews Rating.

The accuracy graph also shows the combinations' prediction at its maximum as per the results. Figure 19 displays the results.

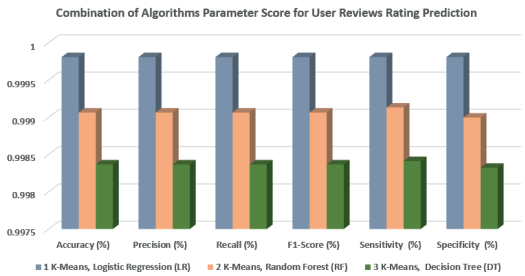

Figure 20: Combination of Algorithms Graph Parameter Based on User Reviews Rating Prediction.

We could limit or alter the precision as per the requirement. For instance, the accuracy, precision, recall, sensitivity, specificity, and f1-Score are at their best correctness, and Figure 20 shows the results.

5. Conclusion

This study focuses on sentiment classification and information retrieval techniques (IRT). It is essential for earlier predictions regarding sentiment classification and IRT. The datasets contain user reviews rating, for which we used multiple algorithms: K-Means Clustering algorithm, Logistic Regression (LR), Random Forest (RF), and Decision Tree (DT) algorithms. Using these algorithms, we get the most efficient outcome, and the accuracy is high due to K-Means, Logistic Regression (LR). We used amazon unlocked mobile reviews and telecom tweets dataset to classify, cluster, and K-Means methodology with the K-Means approach to improving accuracy. As a result of each algorithm being combined with the K-Means, when we combine the K-Means, Logistic Regression (LR), we get an accuracy of 99.98066885752948 %. When we combine the K-Means, and Random Forest (RF), we get an accuracy of 99.90624395901798 %. When we combine the K-Means and the Decision Tree (DT) Algorithm, we get an accuracy of 99.8366518461241 %. Although this study has yielded positive outcomes in sentiment classification and IRT through the use of multiple algorithms, there remains room for more research. Future research can explore the effectiveness of sentiment classification and IRT techniques on diverse datasets, including those in different languages or domains. Investigation into the impact of pre-processing techniques and the use of deep learning methods like CNNs and RNNs can be valuable. Moreover, the effect of different feature selection methods on the accuracy of these algorithms should be explored.

References

Abad-Segura, E., González-Zamar, M.-D., Infante-Moro, J.C., & Ruipérez García, G. (2020). Sustainable management of digital transformation in higher education: Global research trends. *Sustainability, 12*(5), 2107. https://doi.org/10.3390/su12052107

Abualigah, L.M., Khader, A.T., & Hanandeh, E.S. (2018). A hybrid strategy for krill herd algorithm with harmony search algorithm to improve the data clustering? *Intelligent Decision Technologies, 12*(1), 3-14. https://doi.org/10.3233/IDT-170318

Alharbi, A.S.M., & de Doncker, E. (2019). Twitter sentiment analysis with a deep neural network: An enhanced approach using user behavioral information. Cognitive Systems Research, 54, 50-61. https://doi.org/10.1016/j.cogsys.2018.10.001

Arain, M.S., Khan, M.A., & Kalwar, M.A. (2020). Optimization of Target Calculation Method for Leather Skiving and Stamping: Case of Leather Footwear Industry. *International Journal of Business Education and Management Studies (IJBEMS), 7*(1), 15-30. https://www.ijbems.com/doc/IJBEMS-137.pdf

Baig, M.A., Shaikh, S.A., Khatri, K.K., Shaikh, M.A., Khan, M.Z., & Rauf, M.A. (2023). Prediction of Students Performance Level Using Integrated Approach of ML Algorithms. *International Journal of Emerging Technologies in Learning, 18*(1), 216-234. https://doi.org/10.3991/ijet.v18i01.35339

Bansal, J.C., Sharma, H., Jadon, S.S., & Clerc, M. (2014). Spider monkey optimization algorithm for numerical optimization. *Memetic Computing, 6*, 31-47. https://doi.org/10.1007/s12293-013-0128-0

Benavides, L.M.C., Tamayo Arias, J.A., Arango Serna, M.D., Branch Bedoya, J.W., & Burgos, D. (2020). Digital transformation in higher education institutions: A systematic literature review. *Sensors, 20*(11), 3291. https://doi.org/10.3390/s20113291

Boateng, E.Y., Otoo, J., & Abaye, D.A. (2020). Basic tenets of classification algorithms K-nearest-neighbor, support vector machine, random forest and neural network: a review. *Journal of Data Analysis and Information Processing, 8*(4), 341-357. https://doi.org/10.4236/jdaip.2020.84020

Bouazizi, M., & Ohtsuki, T. (2017). A pattern-based approach for multi-class sentiment analysis in Twitter. *IEEE Access, 5*, 20617-20639. https://doi.org/10.1109/ACCESS.2017.2740982

Bouazizi, M., & Ohtsuki, T. (2018). Multi-class sentiment analysis in Twitter: What if classification is not the answer. *IEEE Access, 6*, 64486-64502. https://doi.org/10.1109/ACCESS.2018.2876674

Brownlee, J. (2016). Supervised and Unsupervised Machine Learning Algorithms. *Machine Learning Mastery, 6*(3). https://machinelearningmastery.com/supervised-and-unsupervised-machine-learning-algorithms/

Brownlee, J. (2019). *Machine learning mastery with Weka*. Ebook. Edition, 1(4).

Buriro, M.A., Rahoo, L.A., Nagar, Muhammad Ali Khan; Kalhoro, M., Kalhoro, S., & Halepota, A.A. (2018). Social Media used for promoting the Libraries and Information Resources and services at University Libraries of Sindh Province. *Proceedings of IEEE International Conference on Innovative Research and Development (ICIRD)*. https://doi.org/10.1109/ICIRD.2018.8376293

Channar, P.B., Ahmed, G., Thebo, J.A., Khan, M.A., & Rahoo, L.A. (2023). Factors Of Knowledge Sharing Among Faculty Members In Higher Educational Institutions: An Empirical Study Of The Public Sector. *Journal of Positive School Psychology, 7*(4), 1498-1506. https://journalppw.com/index.php/jpsp/article/view/16622

Chaudhry, A.K., Kalwar, M.A., Khan, M.A., & Shaikh, S.A. (2021). Improving the Efficiency of Small Management Information System by Using VBA. *International Journal of Science and Engineering Investigations, 10*(111), 7-13. http://www.ijsei.com/papers/ijsei-1011121-02.pdf

Chauhan, N.S. (2020). Decision tree algorithm, explained. KDnuggets,[Online]. Available: https://www.kdnuggets.com/2020/01/Decision-Tree-Algorithm-Explained.html .[Accessed 16 April 2021].

Chugh, A., Sharma, V.K., Kumar, S., Nayyar, A., Qureshi, B., Bhatia, M.K., & Jain, C. (2021). Spider monkey crow optimization algorithm with deep learning for sentiment classification and information retrieval. *IEEE Access, 9*, 24249-24262. https://doi.org/10.1109/ACCESS.2021.3055507

Dabbura, I. (2018). K-means clustering: Algorithm, applications, evaluation methods, and drawbacks. *Towards Data Science*.

Datavedas. (2018). Classification Problems. *Datavedas Classification Problems*.

Ducange, P., Fazzolari, M., Petrocchi, M., & Vecchio, M. (2019). An effective Decision Support System for social media listening based on cross-source sentiment analysis models. *Engineering Applications of Artificial Intelligence, 78*, 71-85. https://doi.org/10.1016/j.engappai.2018.10.014

Gao, L., Wang, Y., Li, D., Shao, J., & Song, J. (2017). Real-time social media retrieval with spatial, temporal and social constraints. *Neurocomputing, 253*, 77-88. https://doi.org/10.1016/j.neucom.2016.11.078

Golubic, S., & Marusic, D. (1999). Reviews and inspections-an approach to the improvement of telecom software development process. *Proceedings ConTEL, 99*, 283-290.

Hassan, A.U., Hussain, J., Hussain, M., Sadiq, M., & Lee, S. (2017). Sentiment analysis of social networking sites (SNS) data using machine learning approach for the measurement of depression. *2017 International Conference on Information and Communication Technology Convergence (ICTC)*, 138-140. https://doi.org/10.1109/ICTC.2017.8190959

Injadat, M., Moubayed, A., Nassif, A.B., & Shami, A. (2021). Machine learning towards intelligent systems: applications, challenges, and opportunities. *Artificial Intelligence Review, 54*, 3299-3348. https://doi.org/10.1007/s10462-020-09948-w

Iqbal, F., Hashmi, J.M., Fung, B.C.M., Batool, R., Khattak, A.M., Aleem, S., & Hung, P.C.K. (2019). A hybrid framework for sentiment analysis using genetic algorithm based feature reduction. *IEEE Access, 7*, 14637-14652. https://doi.org/10.1109/ACCESS.2019.2892852

Jianqiang, Z., Xiaolin, G., & Xuejun, Z. (2018). Deep convolution neural networks for twitter sentiment analysis. *IEEE Access, 6*, 23253-23260. https://doi.org/10.1109/ACCESS.2017.2776930

Kaggle. (2023). Amazon Reviews: Unlocked Mobile Phones. https://www.kaggle.com/datasets/PromptCloudHQ/amazon-reviews-unlocked-mobile-phones

Kalwar, M.A., & khan. (2020). Optimization of Procurement & Purchase Order Process in Foot Wear Industry by Using VBA in Ms Excel. *International Journal of Business Education and Management Studies (IJBEMS), 6*(1), 213-220. https://ijbems.com/doc/IJBEMS-124.pdf

Kalwar, M.A., & Khan, M.A. (2020a). Increasing performance of footwear stitching line by installation of auto-trim stitching machines. *Journal of Applied Research in Technology & Engineering (JARTE), 1*(1), 31. https://doi.org/10.4995/jarte.2020.13788

Kalwar, M.A., & Khan, M.A. (2020b). Optimization of Procurement & Purchase Order Process in Foot Wear Industry by Using VBA in Ms Excel. *International Journal of Business Education and Management Studies (IJBEMS), 5*(2), 80-100.

Kalwar, M.A., Khan, M.A., Shahzad, M.F., Wadho, M.H., & Marri, H.B. (2022). Development of linear programming model for optimization of product mix and maximization of profit: case of leather industry. *Journal of Applied Research in Technology & Engineering (JARTE), 3*(1), 67-78. https://doi.org/10.4995/jarte.2022.16391

Kalwar, M.A., Marri, H.B., & Khan, M.A. (2021). Performance Improvement of Sale Order Detail Preparation by Using Visual Basic for Applications: A Case Study of Footwear Industry. *International Journal of Business Education and Management Studies (IJBEMS), 3*(1), 1-22. https://ijbems.com/doc/IJBEMS-159.pdf

Kalwar, M.A., Shahzad, M.F., Wadho, M.H., Khan, M.A., & Shaikh, S.A. (2022). Automation of order costing analysis by using Visual Basic for applications in Microsoft Excel. *Journal of Applied Research in Technology & Engineering (JARTE), 3*(1), 29-59. https://doi.org/10.4995/jarte.2022.16390

Kalwar, M.A., Shaikh, S.A., Khan, M.A., & Malik, T.S. (2020). Optimization of Vendor Rate Analysis Report Preparation Method by Using Visual Basic for Applications in Excel (Case Study of Footwear Company of Lahore). *Proceedings of the International Conference on Industrial Engineering and Operations Management* (IEOM, Dhaka, Bangladesh, December 26-27. https://ieomsociety.org/proceedings/2021dhaka/228.pdf

Kalwar, M.A., Wassan, A.N., Phul, Z., & Wadho, M.H., Malik, T.S., Khan, M.A. (2023). Automation of material cost comparative analysis report using VBA Excel: a case of footwear company of Lahore. *Journal of Applied Research in Technology & Engineering (JARTE), 4*(1), 13-23. https://doi.org/10.4995/jarte.2023.18776

Khan, M.A., Kalwar, M.A., & Chaudhry, A.K. (2021). Optimization of material delivery time analysis by using Visual Basic for applications in Excel. *Journal of Applied Research in Technology & Engineering (JARTE), 2*(2), 89. https://doi.org/10.4995/jarte.2021.14786

Khan, M.A., Kalwar, M.A., Malik, A.J., Malik, T.S., & Chaudhry, A.K. (2021). Automation of Supplier Price Evaluation Report in MS Excel by Using Visual Basic for Applications: A Case of Footwear Industry. *International Journal of Science and Engineering Investigations (IJSEI), 10*(113), 49-60. http://www.ijsei.com/papers/ijsei-1011321-08.pdf

Khan, M.Z., Khan, A.A., Laghari, A.A., Shaikh, Z.A., Kaimkhani, M.A., Morkovkin, D., Gavel, O., Shkodinsky, S., Taburov, D., & Makar, S. (2022). Comparative case study: an evaluation of performance computation between support vector machine, K-nearest comparative study: Evaluation of performance computation between support vector component analysis. *Journal of Tianjin University Science and Technology, April*. https://doi.org/10.17605/OSF.IO/HK3SF

Khan, M.Z., Shaikh, S.A., Shaikh, M.A., Khatri, K.K., Mahira Abdul Rauf, Kalhoro, A., & Muhammad, A. (2023). The Performance Analysis of Machine Learning Algorithms for Credit Card Fraud Detection. *International Journal of Online and Biomedical Engineering (IJOE), 19*(03), 82-98. https://doi.org/10.3991/ijoe.v19i03.35331

Khan, M.Z., Zaman, F.U., Adnan, M., Imroz, A., & Rauf, M.A. (2022). Comparative Case Study: An Evaluation of Performance Computation Between SQL And NoSQL Database. *Sindh Journal of Headways in Software Engineering (SJHSE), 01*(02), 14-23.

Kumar, S., Nayyar, A., Nguyen, N.G., & Kumari, R. (2020). Hyperbolic spider monkey optimization algorithm. *Recent Advances in Computer Science and Communications (Formerly: Recent Patents on Computer Science), 13*(1), 35-42. https://doi.org/10.2174/2213275912666181207155334

Kumar, S., Sharma, B., Sharma, V.K., & Poonia, R.C. (2021). Automated soil prediction using bag-of-features and chaotic spider monkey optimization algorithm. *Evolutionary Intelligence, 14*, 293-304. https://doi.org/10.1007/s12065-018-0186-9

Kumar, S., Sharma, B., Sharma, V.K., Sharma, H., & Bansal, J.C. (2020). Plant leaf disease identification using exponential spider monkey optimization. *Sustainable Computing: Informatics and Systems, 28*, 100283. https://doi.org/10.1016/j.suscom.2018.10.004

Li, L., Xu, Q., Gan, T., Tan, C., & Lim, J.-H. (2017). A probabilistic model of social working memory for information retrieval in social interactions. *IEEE Transactions on Cybernetics, 48*(5), 1540-1552. https://doi.org/10.1109/TCYB.2017.2706027

Mansour, S. (2018). Social media analysis of user's responses to terrorism using sentiment analysis and text mining. *Procedia Computer Science, 140*, 95-103. https://doi.org/10.1016/j.procs.2018.10.297

Mata-Rivera, F., Torres-Ruiz, M., Guzman, G., Moreno-Ibarra, M., & Quintero, R. (2015). A collaborative learning approach for geographic information retrieval based on social networks. *Computers in Human Behavior, 51*, 829-842. https://doi.org/10.1016/j.chb.2014.11.069

Mataoui, M., Sebbak, F., Benhammadi, F., & Bey, K.B. (2015). Query expansion in XML information retrieval: A new approach for terms selection. *2015 6th International Conference on Modeling, Simulation, and Applied Optimization (ICMSAO),* 1-4. https://doi.org/10.1109/ICMSAO.2015.7152208

Matt, C., Hess, T., & Benlian, A. (2015). Digital transformation strategies. *Business & Information Systems Engineering, 57*, 339-343. https://doi.org/10.1007/s12599-015-0401-5

Mbaabu, O. (2020). Introduction to random forest in machine learning. Berreskuratua-(e) Tik Https://Www. Section. Io/Engineering-Education/Introduction-to-Random-Forest-in-Machine-Learning.

Memon, M., Khan, M.A., & Rahoo, L.A. (2020). Usage and Availability of Information and Communication Technology Applications Facilities at Central Library. *International Research Journal in Computer Science and Technology (IRJCST), 1*(1), 86-92. http://irjcst.com/index.php/irjcst/article/view/7/6

Munjal, P., Kumar, L., Kumar, S., & Banati, H. (2019). Evidence of Ostwald Ripening in opinion driven dynamics of mutually competitive social networks. *Physica A: Statistical Mechanics and Its Applications, 522*, 182-194. https://doi.org/10.1016/j.physa.2019.01.109

Munjal, P., Kumar, S., Kumar, L., & Banati, A. (2017). Opinion dynamics through natural phenomenon of grain growth and population migration. *Hybrid Intelligence for Social Networks*, 161-175. https://doi.org/10.1007/978-3-319-65139-2_7

Munjal, P., Narula, M., Kumar, S., & Banati, H. (2018). Twitter sentiments based suggestive framework to predict trends. *Journal of Statistics and Management Systems, 21*(4), 685-693. https://doi.org/10.1080/09720510.2018.1475079

Nagar, M.A.K., Kalhoro, M., & Kalhoro, S. (2018). Information Seeking Behavior of Research Scholars at MUET Library & Online Information Center, Jamshoro: A Study. *Journal of Library Philosophy and Practice, August,* 1-8.

Nagar, M.A.K., Rahoo, L.A., Rehman, H.A., & Arshad, S. (2018). Education management information systems in the primary schools of sindh a case study of hyderabad division. *2018 IEEE 5th International Conference on Engineering Technologies and Applied Sciences (ICETAS)*, 1-5. https://doi.org/10.1109/ICETAS.2018.8629249

Nitze, I., Schulthess, U., & Asche, H. (2012). Comparison of machine learning algorithms random forest, artificial neural network and support vector machine to maximum likelihood for supervised crop type classification. *Proceedings of the 4th GEOBIA*, Rio de Janeiro, Brazil, 79, 3540.

Pant, A. (2019). Introduction to logistic regression. Average. *Towards Data Science*.

Rahoo, L.A., Khan, M.A., Buriro, M.A., Baladi, Z.H., & Abbasi, M.S. (2020). Evaluation of Information Services from the Perspective of Faculties and Evaluation of Information Services from the Perspective of Faculties and Students of Mehran University Engineering and Technology, Jamshoro Pakistan. *International Journal of Disaster Recovery and Business Continuity, 11*(1), 1526-1538. http://sersc.org/journals/index.php/IJDRBC/article/view/20339

Rahoo, L.A., Nagar, M.A.K., & Bhutto, A. (2019). The Use of Information Retrieval Tools by the Postgraduate Students of Higher Educational Institutes of Pakistan. *Asian Journal of Contemporary Education, 3*(1), 59-64. https://doi.org/10.18488/journal.137.2019.31.59.64

Reis, I., Baron, D., & Shahaf, S. (2018). Probabilistic random forest: A machine learning algorithm for noisy data sets. *The Astronomical Journal, 157*(1), 16. https://doi.org/10.3847/1538-3881/aaf101

Reno, U. (2023). *Intelligent Systems*. Department of Computer Science & Engineering, University of Nevada, Reno, USA. https://www.unr.edu/cse/undergraduates/prospective-students/what-are-intelligent-systems

Riverside, U. (2023). *Intelligent Systems*. Department of Electrical and Computer Engineering, University of California, Riverside, USA. https://www.ece.ucr.edu/research/intelligentsystems

Sarker, I.H. (2021). Machine learning: Algorithms, real-world applications and research directions. *SN Computer Science, 2*(3), 160. https://doi.org/10.1007/s42979-021-00592-x

Schott, M. (2019). Random forest algorithm for machine learning. Medium. Com. https://medium.com/capital-one-tech/random-forest-algorithm-for-machine-learning-C4b2c8cc9feb (Erişim 4 Ocak 2021).

Schütze, H., Manning, C.D., & Raghavan, P. (2008). *Introduction to information retrieval* (Vol. 39). Cambridge University Press Cambridge. https://doi.org/10.1017/CBO9780511809071

Shah, I., El Affendi, M., & Qureshi, B. (2020). SRide: An online system for multi-hop ridesharing. *Sustainability, 12*(22), 9633. https://doi.org/10.3390/su12229633

Sharma, A., Sharma, A., Panigrahi, B.K., Kiran, D., & Kumar, R. (2016). Ageist spider monkey optimization algorithm. *Swarm and Evolutionary Computation, 28*, 58-77. https://doi.org/10.1016/j.swevo.2016.01.002

Sheldon, R., & Wigmore, I. (2023). Intelligent System. *Techtarget Network*. https://www.techtarget.com/whatis/definition/intelligent-system

Singhal, A. (2001). Modern information retrieval: A brief overview. *IEEE Data Eng. Bull., 24*(4), 35-43.

Tess, P.A. (2013). The role of social media in higher education classes (real and virtual)-A literature review. *Computers in Human Behavior, 29*(5), A60-A68. https://doi.org/10.1016/j.chb.2012.12.032

Tutorialspoint. (2023). *Classification Algorithms - Random Forest. Machine Learning with Python*, Tutorialspoint. Classification Algorithms - Random Forest

Vial, G. (2019). Understanding digital transformation: A review and a research agenda. *The Journal of Strategic Information Systems, 28*(2), 118-144. https://doi.org/10.1016/j.jsis.2019.01.003

Virmani, C., Juneja, D., & Pillai, A. (2018). Design of query processing system to retrieve information from social network using NLP. *KSII Transactions on Internet and Information Systems (TIIS), 12*(3), 1168-1188. https://doi.org/10.3837/tiis.2018.03.011

Zaman, F.U., Khuhro, M.A., Kumar, K., Mirbahar, N., Khan, Z., & Kalhoro, A. (2021). Comparative Case Study Difference Between Azure Cloud SQL and Mongo Atlas MongoDB NoSQL Database. *International Journal of Emerging Trends in Engineering Research, 9*(7), 999-1002. https://doi.org/10.30534/ijeter/2021/26972021

Zhang, L., Tan, J., Han, D., & Zhu, H. (2017). From machine learning to deep learning: progress in machine intelligence for rational drug discovery. *Drug Discovery Today, 22*(11), 1680-1685. https://doi.org/10.1016/j.drudis.2017.08.010