Vol-08 Issue 05, May -2024 ISSN: 2456-9348
Impact Factor: 7.936

iJETRM
International Journal of Engineering Technology Research & Management

A SYSTEMATIC EXAMINATION OF MACHINE LEARNING ALGORITHMS AND THEIRREAL-WORLD IMPACT

Ms. Helaria Maria
Assistant Professor, New Horizon College, Bangalore

ABSTRACT:
Machine Learning (ML) is an examination of manipulating a structure to act along with learning by supporting with information and update the extreme point of learning in the framework of familiarity snatched at the fitting time of time. This enables systems to upgrade freely after some time ML is a field of programming and a structure that uses computable techniques to empower PC structures to learn with data. ML has been transformed into the placed stock move towards to managing most issues of AI. ML has been working quietly outside of anyone's ability to see for an extended time frame. This paper provides a skeleton examples, future and genuine disillusions of ML and besides covers diverse ML algorithm and there purposes of premium.

KEYWORD:
ML (Machine Learning); AI (artificial intelligence); SR (speech recognition); NLP (Natural Language Processing); IoT (Internet of Things); PR (Pattern Recognition).

I. INTRODUCTION

The name machine learning was a proposal of Arthur Samuel in year 1959 [1]. ML is a sub- set of AI, in this move towards algorithms train themselves by means of data and related information independently. Emerged from the revision of PR and computational learning theory in artificial intelligence [2] ML explore the learning and building of algorithms that can learn to predict on data [3]. ML is strongly related to computational statistics, which as well targets of making forecast with the assist of computers.

ML is a process used to construct complex coded rules (algorithms) and model so as to lend themselves to forecast in business-related use, this is called as predictive analytics. Present day ML coded rules empower computers to communicate with humans, automatically drive cars, find terrorist suspects. Based on the regulations of birth, the expert systems started in 80s. Expert systems were rapidly ratified by the corporate sector, bringing out interest in ML. 1981

— Gerald Dejong introduces the concept of "Explanation-Based Learning" (EBL), in which a computer analyzes the training data and creates general rules allowing the less important data to be discarded. 1985 — Terry Sejnowski invents NetTalk, which learns to pronounce words in the same way a child would learn to make. The field of ML is relaunched based on the potential estimate of ample data available. Many businesses are advancing to incorporate ML into their operations in order stay ahead in their competition [4]

ADVANCEMENTS IN MACHINE LEARNING

Machine learning (ML) is a cutting-edge technology that holds great potential for businesses. Already in use by nearly every Fortune 500 company, ML has enabled organizations to operate more efficiently and generate significant profits. With its innovative capabilities, ML offers promising opportunities for enterprises seeking to stay ahead of the competition.

A. Deeper is a personalization utility.

The users can expertise additional skilled recommendations supported the experiences, the user will need simpler and correct endorsement within the hereafter, the expertise of users can immensely improve.

B. Neural networks running on our mobile devices.

Hand held gadgets might be fit for taking care of ML errands locally, which leaves to manage the cost of a substantial scope of chances for presentations in the portable stage like OB (Object Recognition), FR (Face Recognition) and SR (Speech Recognition). The cell phone may convey the ability to control the machine learning assignments locally, opening an expansive extent of chances for questionable acknowledgment, voice correspondence, confront recognition, and different advancements for versatile stages.

C. Health and fitness

Wellness following wearables in addition to applications be quite famous by the side of this stage. Persons utilize

factors, by associated application towards track their games, exercises, and usual day to day actions. ML can possibly make this a stride further, nonetheless, by giving more point- by-point criticism and tips about a client's action and condition, making wellness trackers more successful.

D. Battery life

This may sound significantly less epic than different conceivable outcomes of machine adapting; however, safeguarding battery life is a standout during the most baffling worries for transportable application clients. Alongside the robotization of structure as set allotment for application, ML could likewise diminish the measure of pointless battery utilization of use.

***C.* Neural systems enhance.**

Neural systems are seemingly the most amazing learning calculations which are available to all at introduction. However, we don't generally see how or why they function. We trust that will change.

***D.* NLP**

ML-based NLP is within such a grieved state, to the point that it can scarcely beat a lead based motor. All things considered, that is a misrepresentation; however, I would state that this subfield of ML is still in its early stages. The principle issue is so as to words encompass distinctive implications in various settings. Algorithms that perceive those unique circumstances and comprehend etymological ideas at a larger amount have not yet been effectively actualized, but rather there's no motivation behind why they can't be.

***E.* Reinforcement learning**

You can accomplish wonderful things with RL. Up until this point, industry ML is generally concerned with supervised learning, picking up bits of knowledge commencing information. The selection of smart specialists will reform numerous ventures later on.

***F.* Quantum Computing**

Machine-learning assignments include issues, for example, controlling and arranging substantial quantities of vectors in high-dimensional spaces. The classical algorithms we as of now use for taking care of such issues require significant investment. Quantum PCs will probably be great at controlling high-dimensional vectors in huge tensor item spaces. It is likely that both the advancement of both managed and unsupervised quantum machine learning calculations will colossally build the quantity of vectors and their measurements exponentially more rapidly than classical algorithms. This will probably bring about a monstrous increment in the speed at which ML algorithms will run.

***G.* Better Unsupervised Algorithms**

UL (Unsupervised learning) happens once no names are known within the direction of the learning algorithmic program; it's left alone towards ascertain structure within the data. UL be able to an objective, for example, finding shrouded styles in information, or a technique towards a finish, times referred to as feature learning. It seems that advances in building brighter, the unsupervised learning algorithmic program may speed up and manufacture additional preciseresults.

***H.* Collaborative Learning**

CL (Collaborative learning) is tied in by method of using different computational essentials so they work together so as to generate favored learning which would have accomplished without anyone else. A case of this would use the hubs of an IoT sensor organize, or what is called edge analytics. Through the maturity of IoT, it is likely that widespread quantities of particular elementwill be worn to learn collaboratively from numerous points of view.

L. Cognitive Services

This innovation incorporates pack like APIs and administrations, through which designers can make more discoverable and canny applications. ML API's will enable the designer to present astute highlights, for example, feeling location or discourse, facial and vision acknowledgment and also dialect and discourse understanding into their applications. The eventual fate of this field will be the presentation of profoundly customized registering encounters for all. These are belongings we think can and ought to occur in the ML splendid prospect, yet it is similarly similar to that the presentation of some new obscure troublesome innovation will bring about a fate of which will we could never have anticipated.

iJETRM
International Journal of Engineering Technology Research & Management

LIST OF POPULAR MACHINE LEARNING ALGORITHM

As technology advances, manual tasks are becoming automated, redefining the term "manual." Machine learning algorithms now come in many varieties, enabling computers to perform a range of tasks, from playing chess to performing surgeries and becoming increasingly intelligentand personalized.

Our world is constantly advancing technologically, and by examining past progress in computing, we can make predictions about what the future holds. One notable aspect of this revolution is the democratization of computing tools and techniques, with data scientists developing sophisticated data-crunching machines in recent years that have yielded impressive results.

Machine learning algorithms, designed during these dynamic times, help solve complex real- world problems by being self-modifying and automated, improving themselves over time. Let's explore the various types of machine learning algorithms and how they are categorized.

Algorithm	Description
1. Linear Regression	In this process, a relationship is established between independent and dependent variables by fitting them to a line. This line is known as the regression line and represented by a linear equation Y = a*X +b [22]
2. Logistic Regression	It's used to estimate discrete values, like binaryvalues 0 or 1, from a set of independent variables.[22] It helps to predict the probabilityof an event by fitting data to a logit function.
3. Decision Tree	This is a supervised learning algorithm that is used for classifying problems. We split the population into two or more homogeneous sets based on the most significant attributes or independent variables.[22]
4. SVM Algorithm	In this Algorithm, plot the raw data as points in an n – dimensional space ie, n = no.of feature, then later tied to a particular coordinate, making it easy to classify the data.[22]
5. Naive Bayes	It's a classifier assumes that the presence of a particular feature in a class is unrelated ti thepresence of any other feature.
6. KNN Algorithm	This algorithm can be applied to both classification and regression problems. It stores all variable cases and classifies any new cases by taking a majority vote of its K neighbors.

iJETRM
International Journal of Engineering Technology Research & Management

7. K Means	It's an Unsupervised learning algorithm, data sets are classified into a particular number of clusters in such a way that all the data points within a cluster are homogenous and heterogeneous from the data in other clusters.[22]
8. Random Forest Algorithm	A collective of decision trees is called a Random Forest, to classify a new object based on its attributes, each tree is classified and the tree votes for that class. The forest chooses the classification having the most votes.
9. Dimensionality Reduction Algorithms	Dimensionality reduction algorithm like decision Tree, Factor Analysis, Missing Value ration and Random Forest can help you findrelevant details.
10. Gradient Boosting Algorithm and Adaboosting Algorithm	These are boosting algorithm used whenmassive loads of data must be handled to make predictions with high accuracy.

PREDICTION IN ACCURACY OF ALGORITHM

1. **Linear Regression** is a linear modelling approach to find the relationship between one ormore independent variables (predictors) as X and dependent variables denoted as Y.

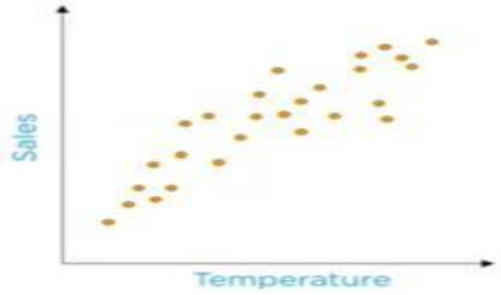


Fig1: Plotting the sales Based temperature.

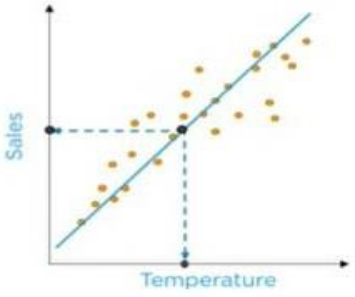


Fig2: Regression line to predict the sales.on

The above figure shows the Temperature is x axis and Sales is y axis which predict sales of ice cream based on temperature. The distances of each point from the lines are classedas regression.[23]

Linear regression is all about finding the best fit line, it can be found out by minimizing the distance between all the data points and the distance to the regression line. The ways to minimize this distance are sum of squared errors, sum of absolute error etc.

iJETRM
International Journal of Engineering Technology Research & Management

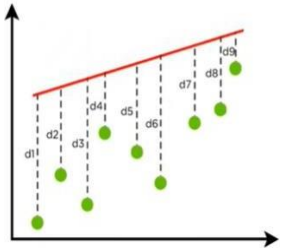


Fig3: Not the best fit line

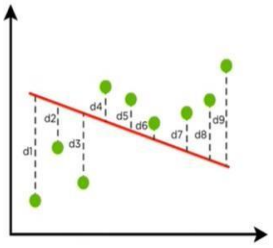


Fig4: Not the best fit line

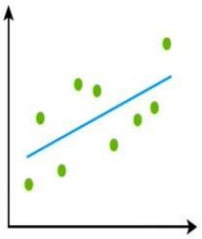


Fig5: Best fit line

Fig 3: Not the best fit line, the distance between the data points and thr line is maximum, Fig4: Not the best fit line, the distance between the data points and the line can still be reduced, To reduce the error we can formulate $\mathbf{D=d1^2+ d2^2 + d3^2+ d4^2+ d5^2+ d6^2+ d7^2+ d8^2+ d9^2}$ we need to square the distances from the point to find the value of D so that the errors is redured, Therefore its proved in [23] Fig5: Best fit line, the distance between the data points and the line is best. Ie, the regressionline (blue) has the least value of D.

2. **Logistic Regression** is a classification algorthim used to predict discrete / categorical values. How does the logistic regression works, if the line is drawn straight to determine the probability, then the perdicated value can excited 0 & 1 its not good one, so the probabilitiy is calculated between 0 & 1 using Sigmoid curve (S) by formulating:

$$\mathbf{P = \frac{1}{1 + e^{-z}}}$$

Here Z is linear equation **m1x+c0**, this is feed to P the value is calculated between 0 & 1 and if Z is negative and much higher value 1/1+0 = 1, else if the Z value is positive and higher value the formula will be P=1. [23]

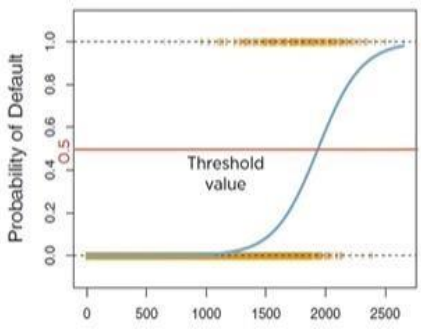


Fig6:Cutoff point at 0.5, anything below it result in 0 and above is 1

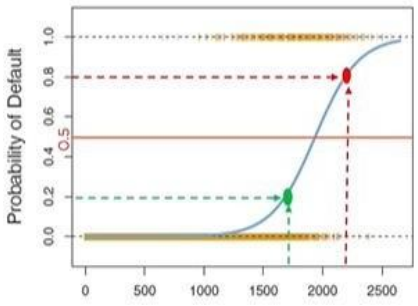


Fig7:Read data point will defult as it is above the threshold value of 0.5 and green data point won't as it is below the threshold value

Example: We can implement by using dataset Socialnetworkads.csv from Kaggle.com, This data set consists of userID, gender, age, estimated salary, purchased. We can predict by using the fieldAge and estimated salary.

iJETRM
International Journal of Engineering Technology Research & Management

	User ID	Gender	Age	EstimatedSalary	Purchased
0	15624510	Male	19	19000	0
1	15810944	Male	35	20000	0
2	15668575	Female	26	43000	0
3	15603246	Female	27	57000	0
4	15804002	Male	19	76000	0
...	...	...	...	...	...
395	15691863	Female	46	41000	1
396	15706071	Male	51	23000	1
397	15654296	Female	50	20000	1
398	15755018	Male	36	33000	0
399	15594041	Female	49	36000	1

iJETRM
International Journal of Engineering Technology Research & Management

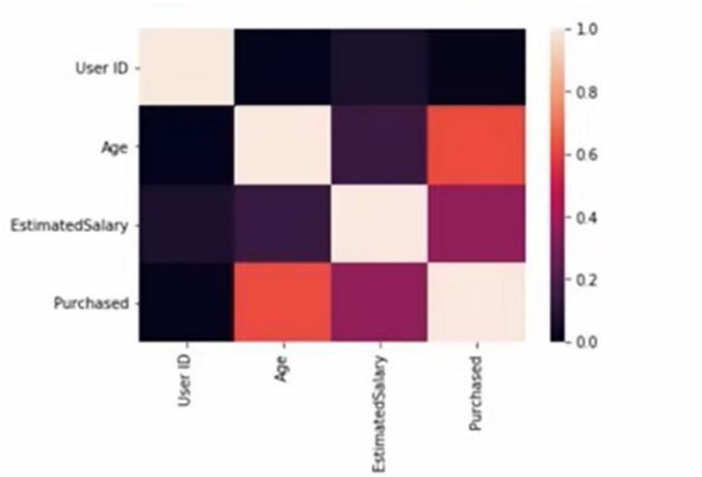
Table 1: SocialNetworkads.CSV dataset [24]

Now we can extract only the independent varibales and dependant varibales.

X = dataset.iloc[:, [2,3]].values # To extract independent variable (2nd and 3rd column ie, age and estimated salary)

Y = dataset.iloc[:,4].values # To extract dependent variable (4th column ie, purchased column) Sns.Heatmap(dataset.corr()) # Heat Map visualising the dataset by drwaing a correlation map

Fig7: Correlation Map



The Drak Color specifies there is no Correlation and lite color means its correlated, there is a sclaing also which denotes the Correlation with scaling 0.0 to 1.0.

Before the training process we need to Splitting the dataset into the training set and test set is used to identify the performes the perdiction correctly or not, whether the model is doing correct or not. Here we split data as 25% & 75% ie,Testsize = 0.25 testdata and remaining 75 for training.

Feature scaling in the data there values will be Varing high and low numbers, here we need normalize the values thatz the work of feature scaling we use standed class **standardscaler() method.**[23]

followed by fitting logistics regression to training dataset , Test set and visualising the Trainingset which result in form of plot.

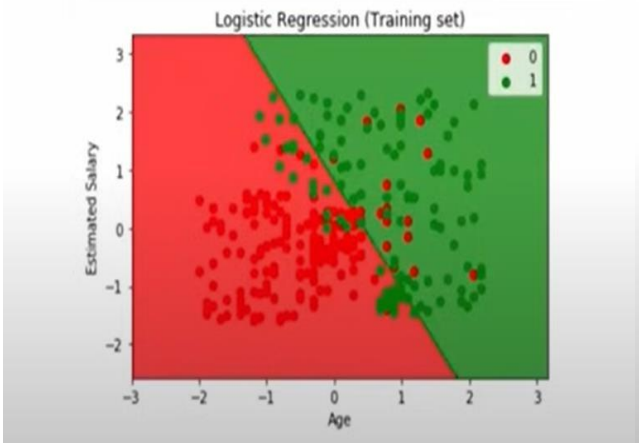


Fig8: Training Set

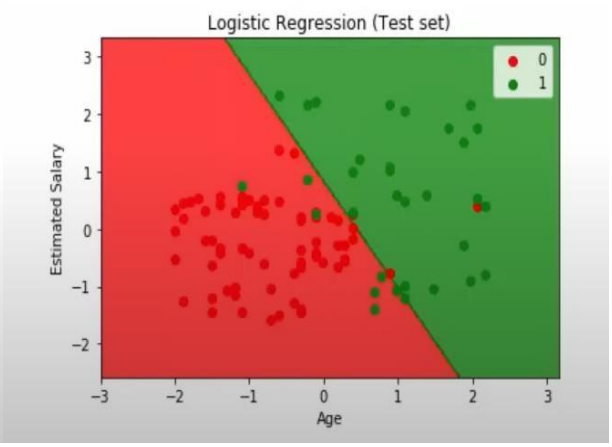


Fig9: Test Set

The classification seen as green is class= 1 , Red is class= 0, Some of the dots are missclassified by red in green area and green in red area. [25]There are Red in red area and green in green area its classified. But this all givies misclassification so We need to Quanlty this accuracy by using Confusion Matrix its readly avaibale methods ie, y_test, y_pred parameters is passed.
Array ([[65, 3], [8, 24]], dtype=int64)
Now The total numbers of values = no of dataset in test data. Ie, 65+3+8+24 = 100 there are total 100 observations in the test dataset. The number in the diagonal is high then there is higher the accuracy.[23]
65+24 = 89
89/100 =0.89%
Therefore 89% of accurayis found.

3. **Decision tree and Random forest** are taken together since they are closely related. Decision Tree is a tree where each node represents a feature (attribute),each link (branch), represents a decision and each leaf represents an outcome (categorical or continous value). its used for classification as well as regression, it can be classificated for multiple classes. Its easy to understand and represent and show how it works.

Example Should the person accept a new job offer?

Its an inverted Tree were the Root node at the top, inner nodes or decision nodes and leaf or terminal nodes are reject or accpect .

- If the salary is greater than 60,000 accpect the offer else reject.
- If the salary is greater than 60,000 Yes then there a decision need to be taken, that thecommute is greater than 1 hour then reject the offer else accpect.
- If the commute is less than 1 hour, then there is a decision need to be then by providingperformance incentives is yes then Accpect the offer else reject the offer

iJETRM
International Journal of Engineering Technology Research & Management

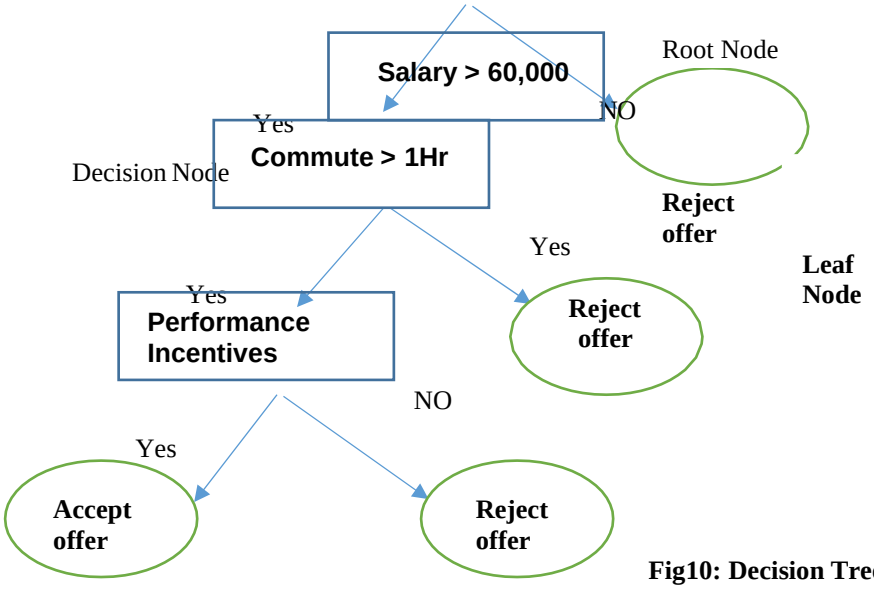


Fig10: Decision Tree Job offer

This all are of different catogories and it below to binary classification. This below to one class accpect offer and another class as reject offer. Lets take an **Example:** We can implement by using dataset kyphosis.csv from Kaggle.com, This data set consists of age, Number, start.

Classification Tree for Kyphosis: does Kyphosis present after a surgery or not.

	Kyphosis	Age	Number	Start
0	absent	71	3	5
1	absent	158	3	14
2	present	128	4	5
3	absent	2	5	1
4	absent	1	4	15

iJETRM
International Journal of Engineering Technology Research & Management

Table2: Dataset kyphosis.Csv

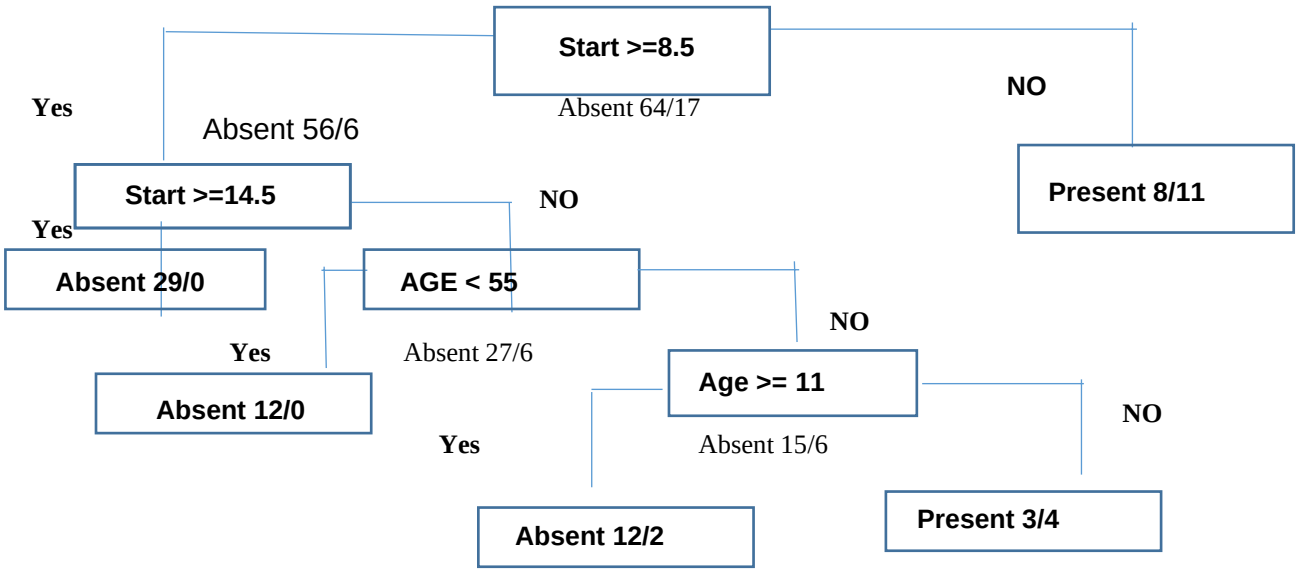


Fig11: Decision Tree kyphosis

iJETRM
International Journal of Engineering Technology Research & Management

The above figure is a Decision tree of Kyphosis. It starts with a criteria if the vertebrae the number on which the surgery is performed, if its greater then or equal to 8.5, furthure analysis is performed else The kyphosis is present. The next criteria if the vertebare is greater than equal to 14.5 is Yes there is absent of Kyphosis. The next criteria is age, if the age is less than 55 age its absent else the next criteria is checked were age is geater than or equal to 111 btween 55 and 111 age, if yesthere is no Kyphosis else there is present of kyphosis.

Lets implement the above logic on decision tree. (I) First we need to load the libraries and then import the dataset and (II) extract the independent and separate the label that is dependent variables by using the python code X=Kyphosis.drop('Kyphosis',axis=1) and Y = Kyphosis['Kyphosis']. (III) visualizing the dataset by using matplotlib. (VI)Split the data into train and test set.(V) Train a decision tree .(IV)Predict the model (VII)Now we classification problem by evaluate the model we use and understand the confusion matrix.

N = 25 total number of observation	Predicted: No	Predicted: Yes
Actual: No	TN = 15	FP = 7
Actual:Yes	FN = 2	TP = 1

Table 3: Model Predication

Accuracy:
(TN+TP)/N

(15+1)/25 = 0.64

Misclassification Rate:
(FP+FN)/N

(7+2)/25 =0.36

Decision Tree Accuracy is 0.64%, Now if we want to improve the performances, here we use Random forest, instead of using one tree we will be using number of trees, Compare he model using random forest Tree, and we take average of all the trees and finally we calssify each tree by its oberveration.

Accuracy:
(TN+TP)/N

(19+0)/25 = 0.76

Misclassification Rate:
(FP+FN)/N

(3+3)/25 =0.24

After applying the Ramdom Forest Tree the performaces is increased and the accuracy is 76%,so we used Ramdon Forest when we are implementing the decision Tree for better accuracy result.

4. **KNN (K Nearest Neighbors) Algorithm** KNN is a classification algorithm generally used to predict categorical values.its very simple and easy to understand. Lets say we have an histroical data with weigth (lbs) and height (ft) and we also have labels then we can classifiy the different classes. Now a new data point is found, that determins which class it belongs too.That point is used as K algorithm this how we train the model.

Let take an example: To find if a new data point is a

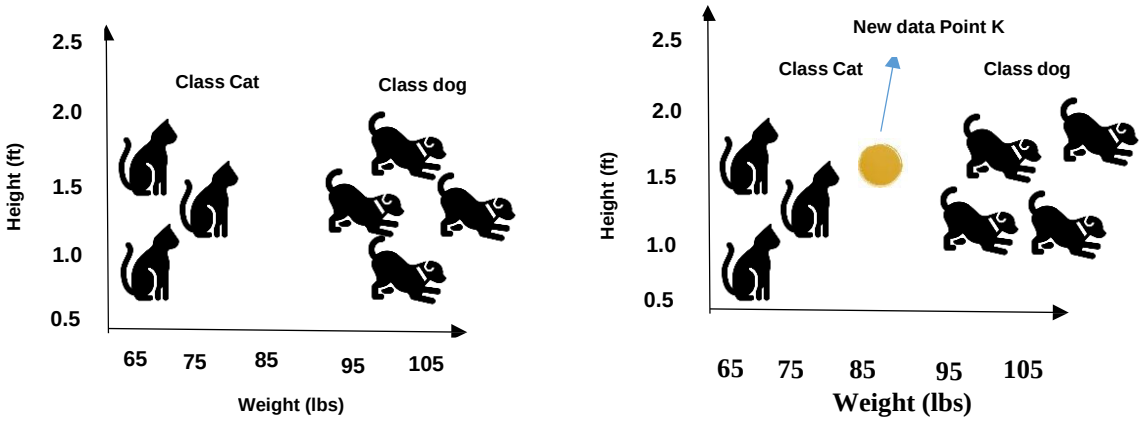


Fig 12: The two classes with the height and weigths

There are two classes i.e. Class dog and Class cat, if the height and weight is in this range its cat,else if the height and the weight is this its dog class, now if we get a point new data set and plot it in the between two classes, how we determine whether it below to cat class or dog class, that were the K comes inn, we need to determine the value of K. using this algorithm we need to determine the value of K by nearest point.

Choose a K will define what class a new data point is assigned to:

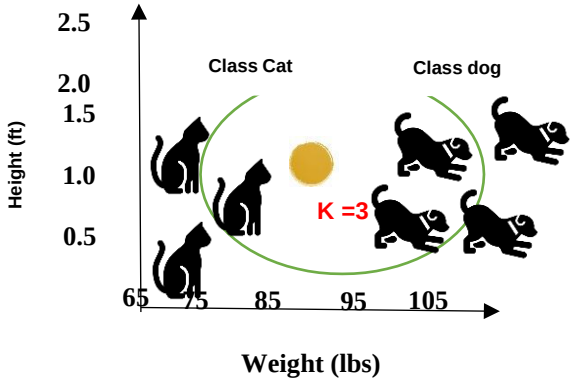


Fig 13:Defining the value of K and choosing the nearest class

If K=3 the nearest three object to this data point, and we find out maximum number of items which class below too, here in the above there are 1 cat and 2 dog the maximum items is dog therefore the new data below to class dog.

Lets implement the above logic on decision tree. (I) First we need to load the libraries and then import the dataset and (II) visualizing the dataset by using matplotlib. (III)Split the data into train and test set.(IV) Fit KNN to Train set (V)Predict the Test Set result (VI) Visualize the Test set Result. (VII) Now we classification problem by evaluate the model we use and understand the confusion matrix.

iJETRM
International Journal of Engineering Technology Research & Management

N = 100 total number of observation	Predicted: No	Predicted: Yes
Actual: No	TN = 64	FP = 4
Actual:Yes	FN = 3	TP = 29

Table 4: Model Predication

Accuracy:
(TN+TP)/N

(64+29)/100 = 0.93

Misclassification Rate:
(FP+FN)/N

(4+3)/100 =0.07

Total number of observation is 100 and KNN Accuracy is 0.93%, and 7 % is only Misclassified its good compared with the above algorithms.

V. RESULTS

The main objective of the study is to explore the different machine learning algorithms to predict the accuracy of each algorithm based on the various input variables which are retained in the model. The classification model was built using several algorithms and each of them using different classification techniques. To approximately evaluate the efficiency and effectiveness of our experiment, we compared few algorithms like, Linear Regression, Logistic Regression, Decision Tree, K-Nearest Neighbor and Random Forest. In Linear Regression the Residuals (errors) are calculated by different means like the mean absolute error MAE, the root mean square RMSE and Mean squared error MSE, this are error and acutual values, the model perdicts MAE is 34%, MSE is 21% and RMSE is 58% which is too less and its better if these values are low the accuracy values are high. In Logistic Regression accuracy is 89% and eror is 11%. In Decision Tree the model perdicts Accuracy is 0.64%, error rate is 36% followed by Random forest model perdicts the accuracy is 76%, error rate is 24%. The KNN model perdicts the accuracy is 93% and error rate is 7%. **Accuracies table shown below.**

	Error Rate	Accuracy
Linear Regression	**0.58%**	**42%**
Logistic Regression	**0.11%**	**89%**
Decision Tree	**0.36%**	**0.64%**
Random Forest	**0. 24%**	**0.76%**
K-Nearest Neighbor	**0.07%**	**0.93%**

Table 5 : Accuracies of Models.

iJETRM
International Journal of Engineering Technology Research & Management

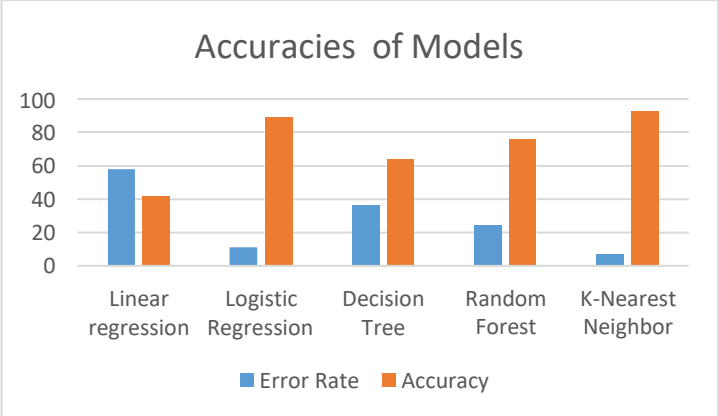


Fig14 : Accuracies of Models

CONCLUSION AND FUTURE SCOPE

In conclusion, Machine learning algorithms have rapidly evolved from theoretical concepts to practical applications with real-world impact. In this systematic examination, we have explored a variety of machine learning algorithms, such as Logistic Regression, Linear Regression, and decision trees. The Proposed automated system emphasizes on prediction of students Feedback rating faculty teaching. The Accuracy of the model is also calculated however, picking the best match model for the student's feedback and exploration the model features will be focused on the future work. The future scope is to build a deep learning classifier and analyze which classifier is more appropriate in carrying out the best classification.

REFERENCES

[1] Samuel, Arthur L. (1988). Computer Games I. Springer, New York, NY. pp. 335365. doi:10.1007/978-1-4613-8716-9_14. ISBN 9781461387183.

[2] http://www.britannica.com/EBchecked/topic/1116194/machine-learning This tertiary source reuses information from other sources but does not name them. Accessed 28/03/2018.

[3] Ron Kohavi; Foster Provost (1998). "Glossary of terms". Machine Learning. 30: 271–274.

[4] http://www.synergicpartners.com/en/espanol-una-breve-historia-del-machine-learning/. Accessed 29/03/2018.

[5] http://www.synergicpartners.com/en/espanol-una-breve-historia-del-machine-learning/. Accessed29/03/2018.

[6] https://www.talend.com/blog/2017/12/04/5-predictions-future-of-machine-learning/. Accessed 30/03/2018.

[7] https://imarticus.org/future-machine-learning-india/. Accessed30/03/2018

[8] https://www.quora.com/What-is-the-future-of-machine-learning-Is-machine-learning-approaching-its-end-Whats-new-in-the-field. Accessed 1/4/2018.

[9] Jaime G. Carbonell, R. S. (1983). Machine Learning: A Historical and Methodological Analysis. Association Samuel, Arthur L. (1988). Computer Games I. Springer, New York, NY. pp. 335–365. doi:10.1007/978-1-4613-8716-9_14. ISBN 9781461387183.

[10] http://www.britannica.com/EBchecked/topic/1116194/machine-learning This tertiary source reuses information from other sources but does not name them. Accessed 28/03/2018.

[11] Ron Kohavi; Foster Provost (1998). "Glossary of terms". Machine Learning. 30: 271–274.

[12]http://www.synergicpartners.com/en/espanol-una-breve-historia-del-machine-learning/. Accessed 29/03/2018.

[13] http://reinforcementlearning.ai-depot.comAccessed:4/04/2018

[14] A Gammerman, V. Vovk, V. Vapnik (1998). "Learning by Transduction." An earlyexplanation of transductive learning.

[15]"A Discussion of Semi-Supervised Learning and Transduction," Chapter 25 of Semi- Supervised Learning, Olivier Chapelle, Bernhard Schölkopf and Alexander Zien, eds. (2006). MIT Press. A discussion of the difference between SSL and transduction.

[16] Chris Fleizach, Satoru Fukushim, “A naive Bayes classifier on 1998 KDD Cup”,Departmentof Computer Science and Engineering University of California, San Diego San Diego.

[17] Domingos, P. and Pazzani, M., 1997. On the optimality of the simple Bayesian classifierunder zero-one loss. Machine learning, 29(2-3), pp.103-130

[18] Rennie, J.D., Shih, L., Teevan, J. and Karger, D.R., 2003. Tackling the poor assumptions ofnaive bayes text classifiers. In ICML (Vol. 3, pp. 616-623).

[19]https://www.linkedin.com/pulse/artificial-neural-networks-advantages-disadvantages-maad- m-mijwel/ Accessed: 10/4/2018.

[20]Tom, M. (1997). Machine Learning. Machine Learning, Tom Mitchell, McGraw Hill,1997:McGraw Hill.

[21] Ryszard S. Michalski, J. G. (1955). Machine Learning: An Artificial Intelligence Approach(Volume I). Morgan Kaufmann

[22] https://towardsdatascience.com/a-tour-of-machine-learning-algorithms-466b8bf75c0a

[23] https://www.slideshare.net/Simplilearn/machine-learning-algorithms-machine-learning- tutorial-data-science-algorithms-simplilearn

[24] https://github.com/PlacementRush/Dataset/blob/main/User_Data.csv

[25] https://clinicaldatascience.io/content/logistic_dataset_python.php