Also as mentioned, Random Forest is used to train our features as well as a developing model for normal and abnormal transactions [1][3]. As here again we basically used Random Forest for minimizing the fraud behaviour in a person's credit card transactions considering the dataset. But Random Forest could not give us accurate and better results on our transactions. So out of the two AdaBoost and KNN, Ada boost had the highest accuracy and better performance.

3. PROPOSED METHODOLOGY-

Our main aim and focus is to basically minimize and reduce fraud transactions and also to classify that both the fraud and non-fraud transactions are present in Dataset, so machine learning algorithms like Random Forest, Ada Boost and KNN are used. These algorithms are compared for the best possible result and outcome.

Process flow diagram for the proposed system is shown in the figure below-

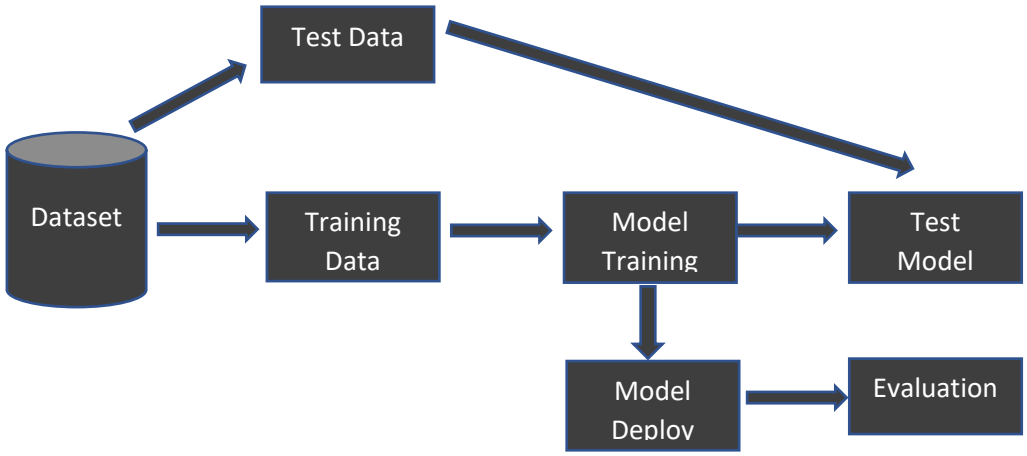

Fig 3: Process flow diagram

While preparing this model training and testing data are considered to train the model using Random Forest, Ada boost and KNN. All the three models are implemented for calculating the accuracy and precision. Finally a comparison is made among these three models for finding the best one[4].

I. Random forest -

It is a popular supervised machine learning algorithm. Random Forest Algorithm can be used for both regression and classification and problems in machine learning [1].

These both have a concept of group (ensemble) learning with them which is a process where we combine many classifiers to resolve complex and difficult problems thus to enhance and increase the accomplishment of the model [5].

This Random Forest is a Classifier that uses many decision trees on various subsets of the specified and declared dataset before averaging the results for improved prediction and dataset precision.

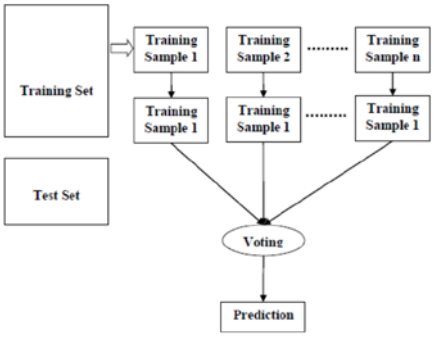

Fig 4: Random forest

Working of Random Forest Algorithm-

1: First, take the Kaggle data set that will be used to develop the credit card fraud detection system, and choose a random K data points from the training set.

2: Create decision trees using the random data to distinguish and categorise fraud from non-fraud cases.

3: Divide the decision tree's nodes by establishing a root node, distinguishing fraud and non-fraud.

4: Now that the majority vote has been completed, we will plot a heat map with a value of 1 for fraud cases and 0 for non-fraud cases.

5: Finding the developed model's accuracy, performance, and precision in both fraud and non-fraud scenarios.

II) Adaboost Algorithm-

Raising, Uplifting and finally Boosting is a Grouping modelling technique that tries to develop a powerful classifier from the amount of weak classifiers. This is done by constructing and building a model with the usage of weak model in that sequence [2].
At first we construct a model coming from the trained data. And then secondly we again construct a model which makes an effort and attempts to solve the errors in the first model.

This Process continues and the models are keep on adding till we predict the complete full data set correctly else the maximum amount of models are added.

The Adaboost algorithm is the initial real desired successful Algorithm for Boosting which has been grown and evolved for the motive of binary classification [6].
Also Adaboost algorithm is a small used for Adaptive Boosting Technique and is a very much well liked and favoured Boosting Method that merges many weak classifiers into strong classifiers.

Basically the Ada Boost algorithm works as a powerful classifier that works well on basic and complex problems.

Working of Ada Boost Algorithm-

1: Firstly take the credit card data set from Kaggle which is trained and then select random data from it.

2: With the random sample data create decision trees for classification of fraud and non-fraud cases**.**
3: The root node of a decision tree can be formed by splitting the node based on high information gain, reclassifying it as fraud and non-fraud cases, and then starting a new decision tree from there.
4: Calculating the performance, error and updating of weights into fraud and non-fraud cases classification.
5: By performing majority vote here and getting the output of decision trees gives the output of non-fraud cases.
6: Decision Trees gives the output of value 1 for fraud case.
7: In the Final Step accuracy, precision and performance for fraud and non-fraud cases are calculated.

III) KNN (K-nearest neighbour's algorithm)

This Algorithm is one of the single and easiest Machine Learning algorithm in which we have a basis of Supervised Learning Technique.

The KNN Algorithm which is there it keeps all the required and accessible data with it and then later it classifies it into a fresh new data point on the basis of the resemblance [7].

Here it gives a meaning that whenever latest data comes in it can be effortlessly classified into a proper suited categorization with the use of our KNN Algorithm Technique.

This KNN Algorithm is also applied for Regression and Classifying but mainly it is used for Classification issues and difficulties.

The KNN Algorithm behaves as a Non-parametric algorithm. Here it means that it is not making any hypothetical or supposition on the fundamental underlying data. Because it doesn't immediately understand or learn from the training set but instead saves and stores the dataset until the scheduled and actual time of classification, hence this algorithm is also known as an idle sluggish, primarily lazy learner algorithm.

Here Also the K-Nearest Neighbour Algorithm on the Training stage simply stores the dataset and as soon as the moment it receives the new data it then classifies that particular data into a class which is very much in resemblance to the new data.

Working of KNN Algorithm-

1: Choose the neighbour with the number K.

2: Determine the Euclidean distance between K neighbours.

3: Using the estimated Euclidean distance, select the K closest neighbours.

4: Surrounded By these K neighbours, maintain track of how many data points there are for each class category.

5: Assign the additional data points to the category where the neighbour count is at its highest.

Evaluation and Result-

Data-Set:

The dataset which we have taken for out this Project is extremely important for all our results and accuracy. So this Dataset we have taken from Kaggle which is a European Company.

The name of the Data set is CreditCard.csv the Data Set has a history of all those transactions which has been from October 2014. The Dataset is a combination of fraud and non-fraud cases.

This Dataset is very huge as it has 284804 rows and 31 columns of transaction history.

The difference in seconds between the current transaction and the initial transaction is represented by the **class time**.

The money transaction is represented by the **Class Account**.

Another Important feature is that the Heat Map which we have plotted shows a Value 1 for Fraud and Value 0 for Non-Fraud which makes our calculation easier.

Evaluation Criteria:

For comparing our these above three algorithms of machine learning we must calculate accuracy, precision and performance of all these.

The Confusion matrix of 2*2 is plotted which evaluates precision and error. There are four outputs in the matrix: FNR, TPR, FPR, and TNR.

Outputs of Confusion Matrix:

1. True positive (TPR): Values that are actually positive and predicated positive.

2. True negative (TNR): Values that are actually negative and predicted negative.

3. False positive (FPR): Values that are actually negative but predicted positive.

4. False negative (FNR): Values that are actually positive but predicted negative.

The Receiver Operating Characteristics (ROC) curve is plotted by TPR vs FPR. Where TPR being the vertical axis and FPR is Horizontal axis. The Graph under ROC Curve is called AUC (area under curve).

Result Analysis:

The dataset which is used gives different outputs and has a difference in performance also with the three algorithms which are random forest, isolation forest, AdaBoost and KNN respectively.

ROC AUC Curve and Confusion Matrix is plotted for these algorithms.

Accuracy Score: 0.9972964432428637
Classification Report:

	precision	recall	f1-score	support
0	1.00	1.00	1.00	28432
1	0.22	0.22	0.22	49

Output of random Forest

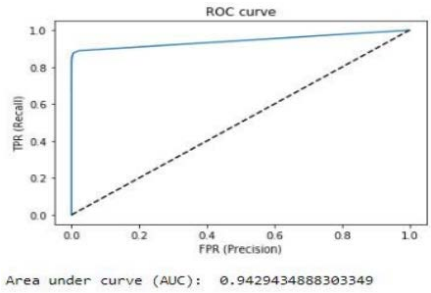

Fig 5: ROC Curve for Random Forest

Now we are putting the output and result of **AdaBoost Algorithm** which is giving very good and best accuracy with performance.

Accuracy on the testing set: 0.999180737616053

Accuracy = 0.9991807377616053
Classification Report:

	precision	recall	f1-score	support
0	**0.99940**	**0.99977**	**0.99958**	**93825**
1	**0.83464**	**0.65432**	**0.73356**	**162**

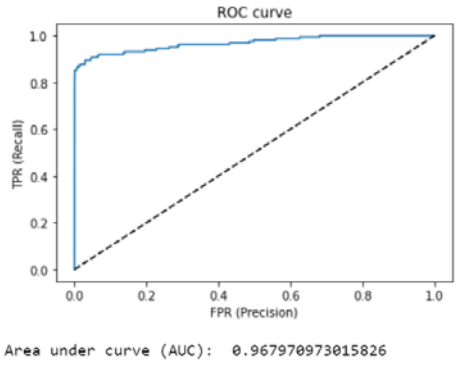

Fig 6: ROC Curve for Ada boost

The Output of Confusion Matrix for Ada Boost

Confusion Matrix on train data:
[190461 29]
[123 207]
Accuracy on the training set: 0.9992034377947803

Confusion Matrix on test data:
[93807 18]
[64 98]
Accuracy on the testing set: 0.9991275316674736
Here applying the results of third algorithm which is KNN (K Nearest Neighbour Algorithm)

Accuracy on testing set: 0.9885487629732

Recall Score of KNN 0.8333333333333334

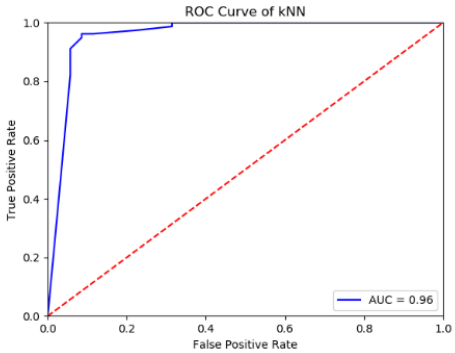

Fig 7: ROC Curve for KNN

From Here on the Comparison of all the Three Algorithms that is **Random Forest, AdaBoost** and **KNN** We find out that:

Random Forest's accuracy is approximately identical to Ada Boost's with a small difference, however KNN's accuracy is lower.

Additionally, the Ada Boost algorithm outperforms the random forest algorithm for fraud cases in terms of precision, Recall and F1-Score.

Hence, in the result analysis we conclude that **The AdaBoost Algorithm** is serving its best with its accuracy and performance for achieving our target of detecting frauds.

4. CONCLUSION:

Apart from many Algorithms in Machine Learning for detecting fraud transactions but we cannot finally confirm and determine the ultimate solution to this problem. We say that we can minimize the fraud behaviour and transactions to some extent with these algorithms but cannot reduce or prevent it completely. Also concluding here that Adaboost algorithm serves best for our main target to achieve which is credit card fraud detection.

5. FUTURE SCOPE:

From the whole analysis and implementation of the Project we have found out that we cannot completely rely on machine learning algorithms for best accuracy and performance as to some or the other extent the results are not correct and not satisfactory as in the case of **Random Forest, KNN and Ada Boost** also to some extent.

So here we would say that we can still use many **deep learning algorithms, AI (Artificial Intelligence) also neural networks** for detecting credit card frauds in the near Future.

REFERENCES:

1. Suharjito Adi Saputra, 'Fraud Detection using Machine Learning in e-Commerce', International Journal of Advanced Computer Science and Applications, Vol. 10, No. 9, 2019.

2. Prashasti Kanikar,Heta Naik, 'Credit card Fraud Detection based on Machine Learning Algorithms, International Journal of Computer Applications' Volume 182 – No. 44, March 2019.

5. Saad Yunus Sait ,Navanshu Khare , 'Credit Card Fraud Detection Using Machine Learning Models and Collating Machine Learning Models', International Journal of Pure and Applied Mathematics Volume 118 No. 20 2018.

7. Abhimanyu ,Roy ,'Deep learning detecting fraud in credit card transactions, 2018 Systems and Information Engineering Design Symposium', IEEE, 2018.

8. Kavya Monisha ,Sahayasakila.V, D., Aishwarya, Sikhakolli Venkata visalakshisesh sai Yasaswi: 'Credit Card Fraud Detection System using Smote Technique and Whale Optimization Algorithm', International Journal of Engineering and Advanced Technology, Volume-8 Issue-5, June 2019.

10. Sarika Jain, Namrata Tiwari, Yashvi Jain, Shripriya Dubey, 'A Comparative Analysis of Various Credit Card Fraud Detection Techniques', International Journal of Recent Technology and Engineering , Volume-7 Issue-5S2, January 2019.

11. Yunyun Zhang,Yong Fang,Cheng Huang, 'Credit Card Fraud Detection Based on Machine Learning', Computers, Materials & Continua CMC, vol.61, no.1, pp.185-195, 2019.

12. Dr. Ajeet Chikkamannur, Kaithekuzhical Leena Kurien, 'Detection And Prediction Of Credit Card Fraud Transactions Using Machine Learning', International Journal Of Engineering Sciences & Research Technology

13. Duman ,Sahin Y. 'E.,Detecting Credit Crad Fraud by Decision Trees and Support Vector Machines‖', International Multi-Conference Of Engineers and Computer Scientists,Mar 16-18,Hong Kong,Vol.1.

14. Jypti Guru, Sai Kiran, Rishabh Kumar, Deepak Katariya, Naveen Kumar, 'Credit card fraud detection using Naïve Bayes model based and KNN classifier‖', Int. Journal of Adv. Research , Ideas and Innovations in Technology,vol.4,2018.

15. D. Aouada,A. C. Bahnsen, and B. Ottersten, 'Example dependent cost-sensitive decision trees', Expert Syst. Appl., vol. 42, no. 19, 2015

16. Kuldeep, Randhawa, 'Credit Card Fraud Detection Using AdaBoost and Majority Voting', IEEE Access, vol. 6, 2018.

17. Shiyang, Xuan, 'Random Forest for Credit Card Fraud Detection', 2018 IEEE 15th International Conference on Networking, Sensing and Control, 2018.

18. John O, Awoyemi, 'Credit Card Fraud Detection Using Machine Learning Techniques: A Comparative Analysis', 2017 International Conference on Computing Networking and Informatics, 2017.

19. V. Gnaneswar, Sailusha, Ruttala, R. Ramesh, and G. Ramakoteswara Rao, 'Credit Card Fraud Detection Using Machine Learning', In 2020 4th International Conference on Intelligent Computing and Control Systems.IEEE, 2020.

20. Dejan,Varmedja, Srdjan Sladojevic, Mirjana Karanovic, Marko Arsenovic, and Andras Anderla, 'Credit card fraud detection-machine learning methods', In 2019 18th International Symposium INFOTEH-JAHORINA, IEEE, 2019.