CREDIT CARD FRAUD DETECTION USING MACHINE LEARNING

Patel Krisha Nilesh[1], Azeem Azad Patel[2], Pranjal Ambadas Bansode[3],Mr. Prashant Rewagad[4]

[1] Student, Computer Department MET Bhujbal Knowledge,MH,India
[2] Student, Computer Department MET Bhujbal Knowledge,MH,India
[3] Student, Computer Department MET Bhujbal Knowledge,MH,India
[4]Professor, Computer Department MET Bhujbal Knowledge,MH,India

ABSTRACT

As the world is rapidly moving towards digitization and money transactions are becoming cashless, the use of credit cards has rapidly increased. The usage of credit cards for online and regular purchases is exponentially increasing and so is the fraud related with it. A large number of fraud transactions are made every day.Online transactions have become a significant and crucial aspect of our lives in recent years. It's critical for credit card firms to be able to spot fraudulent credit card transactions so that customers aren't charged for things they didn't buy. The number of fraudulent transactions is rapidly increasing as the frequency of transactions increases. Since credit card is the most popular mode of payment, the number of fraud cases associated with it is also rising.Thus, in order to stop these frauds we need a powerful fraud detection system that detects it in an accurate manner. Machine Learning and its algorithms can be used to solve such issues.In this paper we have explained the concept of frauds related to credit cards.Here we implement different machine learning algorithms on an imbalanced dataset such as Decision Tree,XGBoost,random forest with ensemble classifiers using boosting technique With Credit Card Fraud Detection, this project aims to demonstrate the modelling of a data set using machine learning. Modeling prior credit card transactions with data from those that turned out to be fraudulent is part of the Credit Card Fraud Detection Problem. The model is then used to determine whether or not a new transaction is fraudulent. Our goal is to detect 100% of fraudulent transactions while reducing the number of inaccurate fraud classifications. Credit Card Fraud Detection is an example of a common classification sample. This Project is focused on credit card fraud detection in real world scenarios. Nowadays credit card frauds are drastically increasing in number as compared to earlier times. Criminals are using fake identity and various technologies to trap the users and get the money out of them. Therefore, it is very essential to find a solution to these types of frauds. In this proposed project we designed a model to detect the fraud activity in credit card transactions. This system can provide most of the important features required to detect illegal and illicit transactions. As technology changes constantly, it is becoming difficult to track the behavior and pattern of criminal transactions. To come up with the solution one can make use of technologies with the increase of machine learning, artificial intelligence and other relevant fields of information technology; it becomes feasible to automate this process and to save some of the intensive amounts of labor that is put into detecting credit card fraud. Initially, we will collect the credit card usage data-set by users and classify it as trained and testing dataset using a random,XGBoost, forest algorithm and decision trees. Using this feasible algorithm, we can analyze the larger data-set and user provided current data-set.The results is indicated concerning the best accuracy for Random Forest are unit 98.6% respectively.

Keyword:-- *XGBoost (Extreme Gradient Boosting), Classifier, Features, Fraud, Train, Accuracy,Random Forest,Decision Tree*

1. INTRODUCTION

A credit card is a thin handy plastic card that contains identification information such as a signature or picture, and authorizes the person named on it to charge purchases or services to his account - charges for which he

will be billed periodically. Today, the information on the card is read by automated teller machines (ATMs), store readers, bank and is also used in online internet banking system. They have a unique card number which is of utmost importance. Its security relies on the physical security of the plastic card as well as the privacy of the credit card number. There is a rapid growth in the number of credit card transactions which has led to a substantial rise in fraudulent activities. Credit card fraud is a wide-ranging term for theft and fraud committed using a credit card as a fraudulent source of funds in a given transaction. Generally, Most of the credit card fraud detection systems are based on artificial intelligence, Meta learning and pattern matching. Credit Card Fraud is described as when a person uses another person's credit card for personal gain while the owner and card issuing authorities are uninformed.Credit card fraud is a simple and inviting target. E-commerce and many other online sites have increased the number of payment options available online, raising the risk of online fraud. Due to the rise in fraud rates, researchers began employing various machine learning approaches to detect and analyze online transaction fraud. Credit card fraud is a common occurrence. Without posing any risks, a large sum of money can be withdrawn in a short period of time without the owner's knowledge. Fraudsters strive to make every fraudulent transaction appear legal, making fraud detection a tough and time-consuming task. Credit Card Fraud is described as when a person uses another person's credit card for personal gain while the owner and card issuing authorities are uninformed. Due to the rise and acceleration of E-Commerce, there has been a massive increase in the use of credit cards for online purchasing, resulting in a high number of credit card frauds. The necessity to identify credit card frauds has become increasingly important in the digital age. Fraud detection entails tracking and analyzing the behavior of a variety of users in order to detect or avoid fraudulent activity. We need to grasp the many technologies, algorithms, and types involved in identifying credit card frauds in order to properly detect credit card frauds. An algorithm can tell whether a transaction is fraudulent or not. To detect fraud, they must have access to a dataset as well as information of fraudulent transactions. They categorize all transactions after analyzing the dataset. Fraud detection is tracking the behaviors of large groups of people in order to predict, detect, or minimize unacceptable activity such as fraud, intrusion, or defaulting.To analyze all permitted transactions and report suspect ones, machine learning techniques like Decision Tree,Random Forest,XGBoost are used. Professionals evaluate these reports and call cardholders to establish whether the transaction was legitimate or fraudulent. The investigators submit feedback to the automated system, which is utilized to train and update the algorithm over time in order to improve fraud detection effectiveness.The use of credit cards is predominant in modern day society and credit card fraud has been kept on increasing in recent years. Huge Financial losses have been fraudulent effects on not only merchants and banks but also the individual person who are using the credits. Fraud may also affect the reputation and image of a merchant causing non-financial losses that. For example, if a cardholder is a victim of fraud with a certain company, he may no longer trust their business and choose a competitor. Fraud Detection is the process of monitoring the transaction behavior of a cardholder to detect whether an incoming transaction is authentic and authorized or not otherwise it will be detected as illicit. In a planned system, we are applying the random forest algorithm for classifying the credit card dataset. Random Forest is an associate in the nursing algorithmic program for classification. Hence, it is a collection of decision tree classifiers. The random forest has an advantage over the decision tree as it corrects the habit of over fitting to their training set. A subset of the training set is sampled randomly so that to train each individual tree and then a decision tree is built, each node then splits on a feature designated from a random subset of the complete feature set. Even for large data sets with many features and data instances, training is extremely fast in the random forest and because each tree is trained independently of the others. The Random Forest algorithm has been found to provide a good estimate of the generalization error and to be resistant to overfitting.

1.1 Problem Statement

There are lots of issues that make this procedure tough to implement and one of the biggest problems associated with fraud detection is the lack of both the literature providing experimental results and of real-world data for academic researchers to perform experiments on. The reason behind this is the sensitive financial data associated with the fraud that has to be kept confidential for the purpose of customer’s privacy. Now, here we enumerate different properties a fraud detection system should have in order to generate proper results. The system should be able to handle skewed distributions, since only a very small percentage of all credit card transactions is fraudulent. There should be a proper means to handle the noise. Noise is the errors that is present in the data, for example, incorrect dates. This noise in actual data limits the accuracy of generalization that can be achieved, irrespective of how extensive the training set is. Another problem related to this field is overlapping data. Many transactions may resemble fraudulent transactions when actually they are genuine transactions.

1.2 Objectives:

The objectives are as follows:

1) To develop an application that is cost efficient.
2) Provide solution with least hardware requirement.
3) ML models can be trained to identify patterns and anomalies in credit card transactions that indicate potential fraud. This can help financial institutions flag suspicious transactions and take appropriate action to prevent losses.
4) ML models aim to minimize the number of falsely flagged transactions as fraudulent. By accurately distinguishing between legitimate and fraudulent transactions, the models can reduce the inconvenience for customers while still identifying and preventing fraudulent activity.
5) ML algorithms can continuously analyze and learn from new data to adapt to evolving fraud patterns. This helps in enhancing the security of credit card systems and staying ahead of fraudsters' techniques.
6) ML algorithms can process large volumes of data in real-time, enabling quick detection of fraudulent transactions. Early detection allows financial institutions to take immediate actions, such as contacting customers or blocking transactions, to mitigate potential losses.
7) By accurately detecting and preventing credit card fraud, ML models can help financial institutions save money by reducing the financial losses resulting from fraudulent activity. Additionally, ML techniques can automate and streamline the fraud detection process, reducing manual effort and associated costs.
8) Effective credit card fraud detection using ML techniques helps build customer trust by providing a secure payment environment. When customers feel confident about the security measures in place, they are more likely to continue using credit cards for transactions.

1.3 Motivation:

Our Project main purpose is to making Credit Card Fraud Detection awaring to people from credit card online frauds. the main point of credit card fraud detection system is necessaryto safe our transactions & security. With this system, fraudsters don't have the chance to make multiple transactions on a stolen or counterfeit card before the cardholder is aware of the fraudulent activity. This model is then used to identify whether a new transaction is fraudulent or not. Our aim here is to detect 100% of the fraudulent transactions while minimizing the incorrect fraud classifications.

2. LITERATURE REVIEW

1) Vimala Devi. J et al. To detect counterfeit transactions, three machine-learning algorithms were presented and implemented. There are many measures used to evaluate the performance of classifiers or predictors, such as the Vector Machine, Random Forest, and Decision Tree. These metrics are either prevalence-dependent or prevalence-independent. Furthermore, these techniques are used in credit card fraud detection mechanisms, and the results of these algorithms have been compared[1].
2) Popat and Chaudhary. supervised algorithms were presented Deep learning, Logistic Regression, Nave Bayesian, Support Vector Machine (SVM), Neural Network, Artificial Immune System, K Nearest Neighbour, Data Mining, Decision Tree, Fuzzy logic based System, and Genetic Algorithm are some of the techniques used. We compared machine-learning algorithms to prediction, clustering, and outlier detection[2].
3) Deepa and Akila . For fraud detection, different algorithms like Anomaly Detection Algorithm, K-Nearest Neighbor, Random Forest, K-Means and Decision Tree were used. Based on a given scenario, presented several techniques and predicted the best algorithm to detect deceitful transactions. To predict the fraud result, the system used various rules and algorithms to generate the Fraud score for that certain transaction[3].
4) Kibria and Sevkli. Using the grid search technique, create a deep learning model. The built model's performance is compared to the performance of two other traditional machine-learning algorithms: logistic regression (LR) and support vector machine (SVM). The developed model is applied to the credit card data set and the results are compared to logistic regression and support vector machine models[4].

5) Borse Suhas and Dhotre Machine learning's Naive Bayes classification was used to predict common or fraudulent transactions.The algorithms used are random forest algorithm and the Adaboost algorithm. The results of the two algorithms are based on accuracy, precision, recall, and F1-score. The ROC curve is plotted based on the confusion matrix. The Random Forest and the Adaboost algorithms are compared and the algorithm that has the greatest accuracy, precision, recall, and F1-score is considered as the best algorithm that is used to detect the fraud [5].
6) Asha R B et al. have proposed a deep learning-based method for detecting fraud in credit card transactions. Using machine-learning algorithms such as support vector machine, k-nearest neighbor, and artificial neural network to predict the occurrence of fraud[6].
7) Ensemble learning methods exploit multiple machine learning algorithms to produce weak predictive results based on features extracted through a diversity of projections on data, and fuse results with various voting mechanisms to achieve better performances than that obtained from any constituent algorithm alone [7].
8) Ensemble learning requires combining the strengths of a number of simpler base models to generate a prediction model. It can be broken down into two tasks: developing a population of base learners from the training data, and then combining them to form the composite predictor [8].
9) The Uncertain Case of Credit Card Fraud Detection:Uncertainty is inherent in many real-time event-driven applications. Credit card fraud detection is a typical uncertain domain, where potential fraud incidents must be detected in real time and tagged before the transaction has been accepted or denied. We present extensions to the IBM Proactive Technology Online (PROTON) open source tool to cope with uncertainty. The inclusion of uncertainty aspects impacts all levels of the architecture and logic of an event processing engine. The extensions implemented in PROTON include the addition of new built-in attributes and functions, support for new types of operands, and support for event processing patterns to cope with all these. The new capabilities were implemented as building blocks and basic primitives in the complex event processing programmatic language. This enables implementation of event-driven applications possessing uncertainty aspects from different domains in a generic manner. A first application was devised in the domain of credit card fraud detection. Our preliminary results are encouraging, showing potential benefits that stemfrom incorporating uncertainty aspects to the domain of credit card fraud detection[9]
10) Fraud is any malicious activity that aims to cause financial loss to the other party. As the use of digital money or plastic money even in developing countries is on the rise so is the fraud associated with them. Frauds caused by Credit Cards have costs consumers and banks billions of dollars globally. Even after numerous mechanisms to stop fraud, fraudsters are continuously trying to find new ways and tricks to commit fraud. Thus, in order to stop these frauds we need a powerful fraud detection system which not only detects the fraud but also detects it before it takes place and in an accurate manner. We need to also make our systems learn from the past committed frauds and make them capable of adapting to future new methods of frauds. In this paper we have introduced the concept of frauds related to credit cards and their various types. We have explained various techniques available for a fraud detection system such as Support Vector Machine (SVM), Artificial Neural Networks (ANN), Bayesian Network, K- Nearest Neighbor (KNN), Hidden Markov Model, Fuzzy Logic Based System and Decision Trees. An extensive review is done on the existing and proposed models for credit card fraud detection and has done a comparative study on these techniques on the basis of quantitative measurements such as accuracy, detection rate and false alarm rate. The conclusion of our study explains the drawbacks of existing models and provides a better solution in order to overcome them[10]

3. PROPOSED SYSTEM

The proposed system of credit card fraud detection using Machine Learning (ML) involves various ML algorithms and techniques to identify and prevent fraudulent transactions. Gather a large dataset containing historical credit card transactions, including both legitimate and fraudulent ones.Clean the dataset by removing any irrelevant or duplicate entries and handling missing values or outliers.Extract relevant features from the dataset that can help in distinguishing between legitimate and fraudulent transactions. Examples of such features include transaction amount, location, time, cardholder information, etc.Choose appropriate ML algorithms for fraud detection. Some commonly used algorithms include, decision trees, random forests and XGBoost.Split the preprocessed dataset into training and testing sets. Train the chosen ML models on the training set using labeled data (i.e., transactions labeled as legitimate or fraudulent).Evaluate the trained models' performance using metrics such as accuracy, precision,

recall, and F1-score on the testing set. Select the best-performing model for deployment.Implement the selected model into a real-time credit card fraud detection system. When a new transaction occurs, the system will analyze its features and predict its likelihood of being fraudulent based on the trained model.If a transaction is predicted to be fraudulent, the system can send an alert to the cardholder and/or the financial institution for further investigation or block the transaction automatically.Regularly update the ML model using new data to improve its performance and adapt to emerging fraud patterns.Continuously monitor the system's performance and collect feedback on identified fraudulent transactions. Analyze the detected fraud cases to identify new patterns and adjust the model or add new features accordingly.By implementing a credit card fraud detection system using ML, financial institutions can significantly reduce the risk of fraudulent activities, provide a safer environment for cardholders.

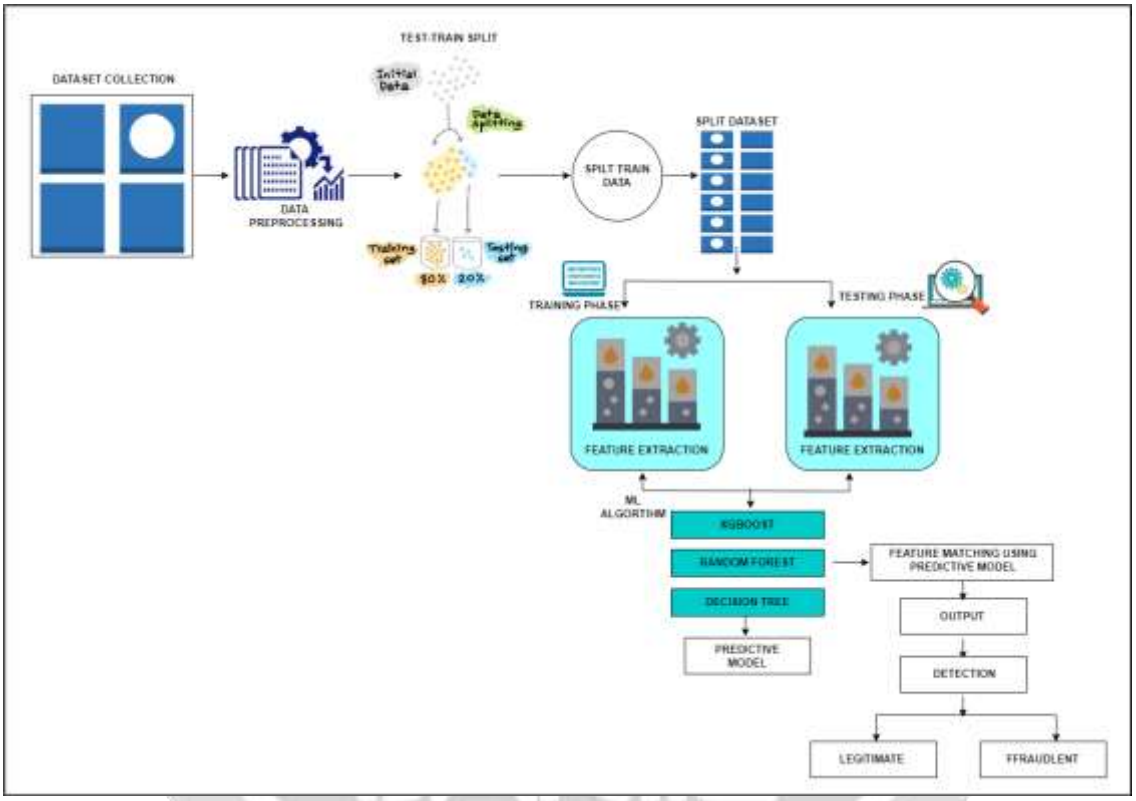

Fig:System Architecture

3.1 Modules

1. Data Collection: Data used in this paper is a set of product reviews collected from credit card transactions records. This step is concerned with selecting the subset of all available data that you will be working with. ML problems start with data preferably, lots of data (examples or observations) for which you already know the target answer. Data for which you already know the target answer is called labelled data
2. Data Pre-processing:Pre-processing is the process of three important and common steps as follows: Formatting: It is the process of putting the data in a legitimate way that it would be suitable to work with. Format of the data files should be formatted according to the need. Most recommended format is .csv files. Cleaning: Data cleaning is a very important procedure in the path of data science as it constitutes the major part of the work. It includes removing missing data and complexity with naming category and so on. For most of the data scientists, Data Cleaning continues of 80% of work. Sampling: This is the technique of analyzing the subsets from whole large datasets, which could provide a better result and help in understanding the behavior and pattern of data in an integrated way

3. Feature extraction: Feature extraction is the process of studying the behavior and pattern of the analyzed data and draw the features for further testing and training. Finally, our models are trained using the Classifier algorithm. We use classify module on Natural Language Toolkit library on Python. We use the labelled dataset gathered. The rest of our labelled data will be used to evaluate the models. Some machine learning algorithms were used to classify pre-processed data. The chosen classifiers were Random forest. These algorithms are very popular in text classification tasks.
4. Evaluation model:Model Evaluation is an essential part of the model development process. It helps to find the best model that represents our data and how well the selected model will work in the future. Evaluating model performance with the data used for training is not acceptable in data science because it can effortlessly generate overoptimistically and over fitted models. To avoid overfitting, evaluation methods such as hold out and cross-validations are used to test to evaluate model performance. The result will be in the visualized form. Representation of classified data in the form of graphs. Accuracy is well-defined as the proportion of precise predictions for the test data. It can be calculated easily by mathematical calculation i.e. dividing the number of correct predictions by the number of total predictions.

3.2 Algorithm

1. **Decision Tree:**

   Decision tree learning is one of the predictive modeling approaches that uses a decision tree (as a predictive model) to go from observations about an item i.e. attribute (represented in the branches) to conclusions about the item's target value i.e. churn or not (represented in the leaves). Tree models where the target variable can take a discrete set of values are called classification trees; in these tree structures, leaves represent class labels and branches represent conjunctions of features that lead to those class labels. Decision trees where the target variable can take continuous values (typically real numbers) are called regression trees.This algorithm splits a data sample into two or more homogeneous sets based on the most significant differentiator in input variables to make a prediction. With each split, a part of a tree is being generated. As a result, a tree with decision nodes and leaf nodes (which are decisions or classifications) is developed. A tree starts from a root node – the best predictor.

2. **Random Forest**:

   The random forest is a classificationalgorithm consisting of many decisionstrees. It uses bagging and featurerandomness when building eachindividual tree to try to create anuncorrelated forest of trees whoseprediction by committee is more accuratethan that of any individual tree.We use Random Forest to predict whether the customer is going to cancel his subscription. Random Forest uses Decision trees for classifying whether the customer is going to cancel his subscription. The random forest consists of a large number of decision trees. A decision tree points to a specific class. A class with more number of votes will be the classifier for a particular customer. Decision trees are sensitive to the data they are trained in. To avoid this, we use Bagging. Bagging is a kind of process where we take a random sample from the dataset for training decision trees.

3. **XGBoost:**

   XGBoost, short for eXtreme Gradient Boosting, is a popular and powerful machine learning algorithm that has been successfully applied to various domains, including fraud detection. It is an ensemble learning algorithm that combines multiple weak prediction models (typically decision trees) to create a strong prediction model.XGBoost (Extreme Gradient Boosting) is a machine learning algorithm that uses an ensemble of weak prediction models, typically decision trees, to build a strong predictive model. It is based on the gradient boosting framework and is known for its scalability, efficiency, and performance.The algorithm works by training decision trees in a sequential manner, where each subsequent tree learns from the mistakes made by the previous trees. The predictions of the individual trees are combined in a additive

manner, with each new tree trying to correct the errors of the previous trees.XGBoost also incorporates regularization techniques, such as shrinkage and subsampling, to prevent overfitting and improve generalization. This helps the algorithm find a good balance between bias and variance, leading to better predictive accuracy.XGBoost has gained popularity and has been widely adopted in various machine learning competitions and real-world applications due to its powerful performance and ability to handle a variety of problem domains.

4. CONCLUSIONS

In conclusion, machine learning algorithms have proven to be effective in detecting credit card fraud. These algorithms are capable of analyzing large amounts of data and identifying patterns that indicate fraudulent activity. By utilizing techniques such as anomaly detection and predictive modeling, machine learning algorithms can accurately identify and prevent instances of credit card fraud.The use of machine learning in credit card fraud detection provides several benefits. Firstly, it helps to minimize losses for both cardholders and card issuers by detecting fraudulent transactions in real-time. By identifying and blocking these transactions, the machine learning models can prevent financial losses and reduce the impact on the victims of fraud.machine learning algorithms continually learn and adapt to new fraud patterns. They can analyze historical data to identify evolving fraud trends and adjust their detection methods accordingly. This adaptive nature ensures that the models remain effective and up-to-date in detecting emerging fraud techniques.it is essential to acknowledge that no method is entirely foolproof, and machine learning algorithms have their limitations. They may produce false positives or false negatives, leading to legitimate transactions being flagged as fraudulent or fraudulent transactions being missed. Human intervention and additional layers of security are still necessary to validate the alerts and further enhance the detection process.

5. REFERENCES

1) N. V. A. Ravikumar, R. S. S. Nuvvula, P. P. Kumar, N. H. Haroon, U. D. Butkar and A. Siddiqui, "Integration of Electric Vehicles, Renewable Energy Sources, and IoT for Sustainable Transportation and Energy Management: A Comprehensive Review and Future Prospects," 2023 12th International Conference on Renewable Energy Research and Applications (ICRERA), Oshawa, ON, Canada, 2023, pp. 505-511, doi: 10.1109/ICRERA59003.2023.10269421.
2) A. K. Bhaga, G. Sudhamsu, S. Sharma, I. S. Abdulrahman, R. Nittala and U. D. Butkar, "Internet Traffic Dynamics in Wireless Sensor Networks," 2023 3rd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), Greater Noida, India, 2023, pp. 1081-1087, doi: 10.1109/ICACITE57410.2023.10182866.
3) Butkar, U. (2015). User Controlling System Using LAN. *Asian Journal of Convergence in Technology*, *2*(1).Uamakant, B., 2017. A Formation of Cloud Data Sharing With Integrity and User Revocation. *International Journal Of Engineering And Computer Science*, *6*(5), p.12.
4) Butkar, U. (2014). A Fuzzy Filtering Rule Based Median Filter For Artifacts Reduction of Compressed Images.
5) Butkar, M. U. D., & Waghmare, M. J. (2023). Hybrid Serial-Parallel Linkage Based six degrees of freedom Advanced robotic manipulator. *Computer Integrated Manufacturing Systems*, *29*(2), 70-82.
6) Butkar, U. (2016). Review On-Efficient Data Transfer for Mobile devices By Using Ad-Hoc Network. *International Journal of Engineering and Computer Science*, *5*(3).

7) Butkar, M. U. D., & Waghmare, M. J. (2023). Novel Energy Storage Material and Topologies of Computerized Controller. *Computer Integrated Manufacturing Systems*, *29*(2), 83-95.
8) Butkar, U. (2014). Synthesis of some (1-(2, 5-dichlorophenyl)-1H-pyrazol-4yl (2-hydroxyphenyl) methanone and 2-(1-(2, 5-dichlrophenyl)-1H-pyrazol-4yl) benzo (d) oxazole. *International Journal of Informative & Futuristic Research (IJIFR)*, *1*(12).
9) Butkar, U. (2014). An execution of intrusion detection system by using generic algorithm.
10) Butkar, M. U. D., & Waghmare, M. J. (2023). Crime Risk Forecasting using Cyber Security and Artificial Intelligent. *Computer Integrated Manufacturing Systems*, *29*(2), 43-57.
11) Navanshu Khare ,Saad Yunus Sait: 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, 825-838 ISSN: 1314-3395.
12) Umakant Butkar, "A Two Stage Crawler for Efficiently Harvesting Web",International Journal Of Advance Research And Innovative Ideas In Education, Vol 2, Issue 3, 2016
13) Umakant Butkar, " Review On- Efficient Data Transfer for Mobile devices By Using Ad-Hoc Network", International Journal of Engineering and Computer Science, vol 5, Issue 3, 2016
14) Umakant Butkar, "User Controlling System Using LAN", Asian Journal of Convergence in Technology, Vol 2, Issue 1, 2015

BIOGRAPHIES (Not Essential)

Patel Krisha Nilesh,
Student,
Computer Department,
MET Bhujbal Knowledge City

Pranjal Ambadas Bansode
Student,
Computer Department,
MET Bhujbal Knowledge City

	Azeem Azad Patel Student, Computer Department, MET Bhujbal Knowledge City
	Mr.Prashant Rewagad Professor, Computer Department, MET Bhujbal Knowledge City Email - prashantr_ioe@bkc.met.edu