SURVEY PAPER ON CHURN PREDICTION FOR BANK CUSTOMER USING MACHINE LEARNING.

Mr. Dattatray Bhimrao Patil, Student, Computer Science Engineering (DS), D.Y.Patil Agriculture and Technical University, Talsande.
Prof. Naresh Ashok Kamble, Asistant Professor, Computer Science and Engineering, D.Y. Patil Agriculture and Technical University, Talsande.

ABSTRACT:
In the banking industry, a customer dropout describes the situation when a customer ceases to use the products and services offered by the bank for a time, then break off the banking relationship. Therefore, in today's highly competitive banking world, customer retention is a determinant. Moreover, customers or even more significant clients increase confidence and word-of-mouth referrals through the loyalty of existing customers, which creates new business opportunities. Because of these facts, banks need to take the crucial step of reducing client attrition. The objective of our research is to analyse bank data and predict which users are most likely to stop using the bank's services and start making payments. For evaluating the data and presenting comparative analysis on numerous evaluation criteria, we use a variety of machine learning methods.
Keywords— Bank customer attrition, Churn prediction, Machine learning.

INTRODUCTION:
Customer attrition, also known as customer churn, is the instance when clients end their relationship with a company or organization. Customer attrition in the banking industry refers to when consumers stop using a particular bank's services or close their accounts. Therefore, understanding and managing customer attrition is imperative for banks to maintain their financial health and protect their brand. Customer attrition can cost banks a huge amount of money and result in loss of revenue for various banking services. Building long-term client relationships is therefore highly valuable to banks. Knowing the trends about attrition, banks can determine clients at risk of leaving and initiate retention strategies. This is a strategy that enhances bank profitability and increases general client lifetime value. Besides this, customer attrition also has a negative impact on the reputation and brand impression of a bank. High attrition rates often indicate deeper issues, such as poor customer service, inefficient procedures, or the lack of competitive services and products. To get past these issues and enhance the overall customer experience, banks must understand and manage client attrition. One of the important tools in analyzing customer attrition is the Data Visualization RShiny app, a web application framework developed using the R programming language. This tool enables users to interact with dashboards and visualizations about churn- related data, allowing for deeper comprehension of the information and making it possible to spot trends and patterns that are associated with customer attrition.

Any business model aspiring to succeed should have a sufficient number of customers, which basically means achieving two major objectives: acquiring new customers and retaining the existing ones. The creation of products and promoting them to relevant markets are essentials in acquiring new customers. Since lost consumers are very unlikely to come back, the second challenge—keeping customers—is crucial for any company strategy to succeed. The key focus in our problem statement is keeping customers and anticipating their trends, which ultimately helps to resolve the issue of customer attrition.

The purpose of this study is to broaden the use of the previously described CRM systems, with an emphasis on determining and forecasting the probability of client attrition. The results of this research can be used in practical situations to help banks identify customer attrition and implement retention strategies. In a separate study, the authors looked into the profit-loss ratio in relation to

when customers stop using products and talked about managing churn to maximize earnings. In addition to offering useful applications for forecasting bank customer attrition, this work serves as a foundation for future research in this area. This study holds potential implications for stakeholders in the Banking industry. Better customer service, tailored banking solutions, and personalized offers are all results of more effective client retention strategies, and they will all improve consumer experiences. Employees may experience improved work conditions and increased job satisfaction by allocating resources and training initiatives aimed at improving customer service. Shareholders might expect better financial performance when customer churn declines and customer lifetime value and profitability rise. Applying research findings can also improve a bank's reputation and brand impression, draw in new customers, and promote sustained business expansion. The study also highlights how important data-driven decision making is, allowing stakeholders to make well-informed choices based on insights from churn analysis and promoting an evidence- based decision-making culture across the sector.

This study adds significantly to our understanding of customer churn analysis and machine learning in the banking industry. The article's first section addresses the crucial problem of data preparation specific to customer churn research in banking by outlining a thorough pretreatment technique that ensures data accuracy and consistency. Second, the study looks closely at several machine learning techniques and assesses how well they predict customer attrition. This comparison study provides insightful information about how well different churn prediction algorithms function in the banking sector. Additionally, the creation of the Data Visualization RShiny app improves the usefulness of churn analysis by providing an easy-to-use tool for presenting churn-related findings.

LITERATURE:
This literature survey study focuses on the key drivers of customer churn, predictive modeling techniques, and strategies for retention in the industry. As such, by analyzing past studies, this particular study aims to identify gaps and emerging trends that can enhance customer retention strategies and also help in developing banking services.

[1] Amuda, K.A., Adeyemo, A.B., 2019. Customers churn prediction in financial institution using artificial neural network
The challenge of customer churn has significant implications for financial institutions concerning revenue, profitability, and long-term relationships with customers. Accurate prediction of the churn will help banks embrace proactive retention strategies to enhance customer satisfaction. This paper investigates using ANNs-an accurate machine learning methodology-to identify the likelihood of customer churn and their behavior in regards to their purchase patterns over time. This study identifies key drivers of churn based on transaction history, account activity, customer demographics, and service usage. Using real-world banking datasets, a predictive model based on ANN is designed to classify customers as at-risk churners or loyal clients. Performance metrics like accuracy, precision, recall, and F1-score for performance evaluation confirm the efficiency of the ANN model in identifying customers who are likely to churn. The study highlights the benefits of ANNs over traditional statistical methods in churn prediction, offering financial institutions a data-driven approach to enhance customer retention strategies and optimize business operations.

[2] Baghla, S., Gupta, G., 2022. Performance evaluation of various classification techniques for customer churn prediction in E-commerce. Microprocess. Microsyst. 94 (Oct.), 104680.
Customer churn prediction is a crucial challenge for e- commerce businesses, as retaining existing customers is more cost-effective than acquiring new ones. In this study, Baghla and Gupta (2022) evaluate the performance of various classification techniques for predicting customer churn in e-commerce. The research compares machine learning algorithms, including decision trees, support vector machines, random forests, and deep learning models, to determine their effectiveness in identifying at-risk customers. Using key customer behavior attributes such as purchase history,

browsing patterns, and engagement metrics, the study assesses model performance based on accuracy, precision, recall, and F1-score. The findings highlight the strengths and limitations of different classifiers, providing insights into the most suitable techniques for churn prediction in the e-commerce domain. The study contributes to the growing field of predictive analytics by offering data-driven recommendations for improving customer retention strategies.

[3] De Lima Lemos, R.A., Silva, T.C., Tabak, B.M., 2022. Propension to customer churn in a financial institution: a machine learning approach. Neural Comput. Appl. 34 (14), 11751–11768.

Customer churn is a critical issue for financial institutions, affecting revenue stability and long-term profitability. In their research, De Lima Lemos et al. (2022) examine the propensity of customer churn using machine learning techniques. The study employs several predictive models, such as tree-based classifiers, support vector machines, and deep learning methods, to compare their performance in predicting customers likely to churn from a financial institution. Based on the data related to customers' transaction information, account activities, and demographic characteristics, this research examines model performance with various evaluation metrics including accuracy, precision, recall, and AUC-ROC. Findings: Insights of factors determining churn in addition to a general illustration of machine learning-based improvements to the strategies used in retaining customers The research further enhances the existing area of financial analytics with its novel approach for proactive management of customer churn through a data-driven framework.

[4] Dias, J., Godinho, P., Torres, P., 2020. Machine learning for customer churn prediction in retail banking. In: International Conference on Computational Science and its Applications. Springer, Berlin, pp. 576–589.

Customer churn prediction is one of the critical challenges for retail banking, as retaining existing customers is essential for sustaining profitability and competitive advantage. In their study, Dias et al. (2020) apply machine learning techniques to predict customer churn in retail banking. The research explores various classification algorithms, including decision trees, support vector machines, random forests, and neural networks, to identify customers at risk of leaving. Utilizing a dataset comprised of customer transaction history, demographic information, and service usage patterns, the research measures model performance on key metrics including accuracy, precision, recall, and F1-score. Results indicate machine learning's predictive capabilities for identifying churn and insight into the factors that most powerfully drive attrition. This study contributes to the financial analytics domain by showing how data-driven approaches can enhance customer retention strategies in retail banking.

[5] Domingos, E., Ojeme, B., Daramola, O., 2021. Experimental analysis of hyperparameters for deep learning-based churn prediction in the banking sector. Comput. Times 9 (3), 34.

Predicting customer churn is a very crucial function for banks in retaining the right clients and remaining profitable. In their study, Domingos et al. (2021) carry out an experimental analysis of hyperparameter tuning for deep learning-based churn prediction in the banking sector. The research examines the effect of various hyperparameters such as learning rate, batch size, activation functions, and network depth on the performance of deep learning models. Using banking customer data, including transaction history, account activity, and demographic features, the study applies neural networks and assesses their predictive accuracy by metrics such as precision, recall, F1-score, and AUC-ROC. Results emphasize the relevance of hyperparameter optimization in developing better churn prediction models and inform optimal configurations of deep learning approaches. This research contributes to financial data science by showing how fine-tuning deep learning models can improve customer retention strategies in banking.science by showing how fine-tuning deep learning models can improve customer retention strategies in banking.

[6] Geiler, L., Affeldt, S., Nadif, M., 2022. An effective strategy for churn prediction and customer profiling. Data Knowl. Eng. 142 (Nov.), 102100.

Customer churn prediction is essential for businesses that intend to retain their clients and maximize

long-term profitability. In this study, Geiler et al. (2022) outline an effective approach to churn prediction and customer profiling using advanced data-driven techniques. The research combines machine learning models with clustering methods to identify at-risk customers with precision while producing detailed customer profiles. This study evaluates various classification algorithms, namely, decision trees, random forests, and neural networks, by basing the entire prediction on transactional data, demographic information, and behavioral patterns. Moreover, customer segmentation techniques are employed to identify key attributes that characterize different churn risk levels. The results of this study highlight how integration of predictive analytics with customer profiling enhances retention. This study provides a robust framework for proactive churn management and personalized marketing in multiple industries and therefore contributes to data-driven decision-making.

[7] Guliyev, H., Tatoglu, F.Y., 2021. Customer churn analysis in banking sector: evidence from explainable machine learning models. J. Appl. Mic. Econ. 1 (2), 85–99.

In the banking field, customer churning is the major issue in this industry due to its real-time impact on profitability and longer-lasting relationships of customers. Utilizing explainable machine learning approaches to analyze customers' churn improves the interpretability and decision making for Guliyev and Tatoğlu, (2021). The article used various forms of predictive models to identify principal reasons for this churning effect such as the use of Gradient Boosting technique, Decision trees, and the use of SHAP values by Shapley Additive Explanation. The study evaluates the performance of the model by accuracy, precision, recall, and AUC-ROC scores based on customer transaction history, demographic characteristics, and engagement measures. The results ensure openness in machine learning decisions while providing insightful information about consumer behavior, focusing on the most important churn predictors. This study advances the field of financial analytics by highlighting the importance of explainability in AI-driven churn prediction and supporting banks.

[8] He, B., Shi, Y., Wan, Q., et al., 2014. Prediction of customer attrition of commercial banks based on SVM model. Procedia Comput. Sci. 31 (Jan.), 423–430.

Customer attrition threatens revenue and long-term competitiveness for commercial banks. In his research study, He et al. (2014) introduce an SVM-based model to predict customer attrition. In this regard, the study makes use of historical banking data including client transactions, account activity, and demographic characteristics to train and test the SVM model. The accuracy, precision, recall, and F1-score metrics are used to measure the performance of the model, showing how well the model identifies at-risk clients. It demonstrates the SVM's advantage in managing complex, high-dimensional data for churn prediction by comparing it with traditional statistical methods. The report offers a data-driven approach to enhance retention strategies in commercial banking as well as key aspects that affect client attrition.

[9] Ho, S.C., Wong, K.C., Yau, Y.K., et al., 2019. A machine learning approach for predicting bank customer behavior in the banking industry. In: Machine Learning and Cognitive Science Applications in Cyber Security. IGI Global, pp. 57–83.

To maximize financial services and enhance client retention, banks must understand and predict consumer behavior. Ho et al. (2019) use machine learning approaches in this study to predict the behavior of bank customers, including transaction patterns, product preferences, and churn propensity. In order to assess consumer data, the study investigates a number of categorization models, including decision trees, support vector machines, and deep learning techniques. The assessment of model performance is made using accuracy, precision, recall, and AUC-ROC scores. The study utilizes demographic data, transaction history, and engagement indicators to assess model performance. The results show how the machine learning works to identify intricate behavioral patterns and improve decision-making in the banking industry. The work provides insights for risk management, client relationship optimization, and tailored marketing of data-driven banking tactics.

[10] Karvana, K.G.M., Yazid, S., Syalim, A., et al., 2019. Customer churn analysis and prediction using data mining models in banking industry. In: 2019 International Workshop on Big Data and Information Security. IEEE, pp. 33–38.
In the banking sector, customer attrition is a critical issue that affects customer relationship management and profitability. Karvana et al. (2019) utilize data mining models to predict and evaluate client attrition in the banking sector. To determine the clients at risk of leaving, the research examines several machine learning techniques, including support vector machines (SVM), decision trees, and random forests. For accuracy, precision, recall, and F1- score, this study evaluates model performance using client transaction history, demographic information, and service consumption trends. Outcomes show the effectiveness of data mining techniques to identify churn trends and increase the accuracy of prediction. This work in banking analytics advances by offering insights into critical churn reasons and suggesting data-driven tactics for risk mitigation and customer retention.

[11] Machado, M.R., Karray, S., 2022. Applying hybrid machine learning algorithms to assess customer risk- adjusted revenue in the financial industry. Electron. Commer. Res. Appl. 56 (Nov.), 101202.
Customer risk-adjusted revenue should be assessed for effective profitability and risk management. Machado and Karray (2022) in their research introduce a hybrid machine learning approach for the assessment of customer value incorporating associated risks. The paper uses multiple machine learning algorithms that are ensemble methods and deep learning techniques to predict the customer revenue adjusted for financial risk factors. Using the data of transaction, credit history, and behavioral attributes, model performance is tested by the researcher using important measures of evaluation including accuracy, precision, and AUC-ROC. The research outcome proves hybrid models are the best to explain intricate financial phenomena and generate superior revenue predictions. This study thus contributes to the financial analytics community by offering a data-driven approach to customer segmentation, risk estimation, and strategic decision-making within the financial domain.

[12] Rahman, M., Kumar, V., 2020. Machine learning based customer churn prediction in banking. In: 2020 4th International Conference on Electronics, Communication and Aerospace Technology. IEEE, pp. 1196–1201.
The banking industry faces a serious problem with customer attrition, which has an impact on both profitability and client retention. Rahman and Kumar (2020) investigate machine learning methods for forecasting banking customer attrition. The study analyzes consumer behavior and identifies consumers who are at danger by using a variety of classification techniques, such as logistic regression, decision trees, random forests, and support vector machines (SVM). Accuracy, precision, recall, and F1-score are the metrics used in the study to assess model performance using financial transaction data, demographic data, and service consumption patterns. The results demonstrate how well machine learning works to identify churn trends and raise prediction accuracy. By helping financial institutions enhance their marketing and risk management initiatives, this study aids in the creation of data-driven customer retention strategies.

PROPOSED WORK:
 The research will be divided into two main components:
  i)Machine Learning Model Development
  ii)Visualization App Development.

DATA COLLECTION AND PREPROCESSING :
Data Source: Use a publicly available banking dataset (e.g., Kaggle) or collaborate with a bank to obtain real- world customer data.

Features: Include demographic data, transaction history, account balances, customer service interactions, and other relevant features.

Preprocessing:

Handle missing values, outliers, and imbalanced data.

Perform feature engineering to create meaningful predictors (e.g., average transaction amount, frequency of complaints).

Normalize or standardize data for ML models.

1. Exploratory Data Analysis (EDA)
   - Perform statistical analysis to understand data distribution and relationships.
   - Use visualizations (e.g., heatmaps, histograms, scatter plots) to identify patterns and correlations.

Identify key factors contributing to churn (e.g., low accounts balance, frequent complaints

MACHINE LEARNING MODEL DEVELOPMENT:

Model Selection: Experiment with multiple ML algorithms:
- Logistic Regression
- Decision Trees
- Random Forest
- Gradient Boosting (XGBoost, LightGBM)
- Neural Networks

Model Training: Split data into training and testing sets (e.g., 80:20 ratio).

Hyper parameter Tuning: Use techniques like Grid Search or Random Search to optimize model performance.

Evaluation Metrics: Use accuracy, precision, recall, F1-score, and ROC-AUC to evaluate model performance.

Feature Importance: Identify the most significant features influencing churn.

VISUALIZATION APP DEVELOPMENT:

Purpose: Create an interactive app for data scientists and managers to explore churn predictions and insights.

Features:

•Dashboard displaying key churn metrics (e.g., churn rate, top predictors).

• Interactive visualizations (e.g., bar charts, pie charts, heat maps).

•Filter options to analyze specific customer segments.

•Real-time predictions using the trained ML model.

Tools: Use Python libraries like Streamlit, Dash, or Shiny for app development.

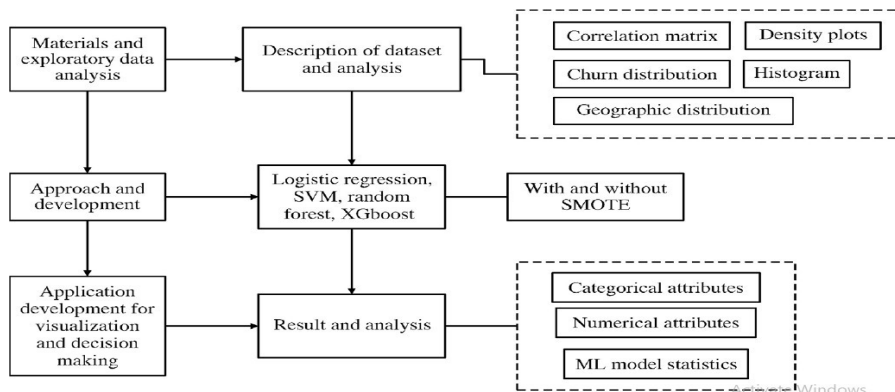**User Interface:** Design a simple and intuitive interface for non-technical users.

Fig.1. Proposed System Architecture

CONCLUSION:

With moderate success, this study helps in forecasting bank client attrition. There is still some room for development, though. The access to large datasets is limited because data on banking are sensitive. More broadly applicable would predictions be, if there were extra data points. Further details would also imply better estimates about the customers by having access to more detailed data. The attributes are more apt for a profile of a customer compared to its recency which is the measurement that documents an activity just before churning. These indicators will be available for prospective researchers that can use the same to make better identification on churn tendencies and track a change in consumer behavior right before a customer leaves. Prospective researchers can also improve our program through other characteristics and data points, the automatic training process of the model, and the development of new models. This way, by changing the patterns and expanding the set of data, it would help bring about a feedback loop within the models, thus ensuring greater accuracy in prediction. They canalso further implement the additional prediction algorithms within the app for visualization with a better outcome on controlling churn and comparative analysis. This would allow the implementation of this app as a consolidated churn management system for multiple companies. Another possible area of research could be the development of localized prediction models that predict turnover for a limited subset of consumers. For example, the datasets can contain different models for clients in different countries. The idea is that, rather than a model intended to fit every scenario, individual models will produce higher accuracy and progressively greater generalizability. Moreover, the study demonstrates that different prediction algorithms produce more impactful forecasts by being better suited for various client categories.

.

REFERENCES:

[1] Ray, P.P. Internet of Things for smart agriculture: Technologies practices and future direction. J. Ambient. Intell. Smart Environ. 2017, 9, 395–420.

[2] Kamienski, C.; Soininen, J.-P.; Taumberger, M.; Dantas, R.; Toscano, A.; Salmon Cinotti, T.; Filev Maia, R.; Torre Neto, A. Smart Water Management Platform: IoT-Based Precision Irrigation for Agriculture. Sensors 2019, 19, 276.

[3] Ojha, T.; Misra, S.; Raghuwanshi, N.S. Wireless sensor networks for agriculture: The state-of-the-art in practice and future challenges. Comput. Electron. Agric. 2015, 118, 66–84.

[4] Vijayan, T.; Sangeetha, M.; Kumaravel, A.; Karthik, B. Feature selection for Simple Color Histogram Filter based on Retinal Fundus Images for Diabetic Retinopathy recognition. IETE J. Res. 2020, 1–8.

[5] Lavanya, G.; Rani, C.; GaneshKumar, P. An automated low cost IoT based Fertilizer Intimation System for smart agriculture. Sustain. Comput. Inform. Syst. 2020, 28, 100300.

[6] Sivakumar, M.; Renuka, P.; Chitra, P.; Karthikeyan, S. IoT incorporated deep learning model combined with SmartBin technology for real-time solid waste management. Comput. Intell. 2021.

[7] Katarya, R.; Raturi, A.; Mehndiratta, A.; Thapper, A. Impact of Machine Learning Techniques in Precision Agriculture. In Proceedings of the 2020, 3rd International Conference on Emerging Technologies in Computer Engineering: Machine Learning and Internet of Things, ICETCE, Jaipur, India, 7–8 February 2020; pp. 1–6.

[8] Anitha, P.; Chakravarthy, T. Agricultural Crop Yield Prediction using Artificial Neural Network with Feed Forward Algorithm. Int. J. Comput. Sci. Eng. 2018, 6, 178–181.

[9] Anand, R.; Karthiga, R.D.; Jeevitha, T.; Mithra, J.L.; Yuvaraj, S. Blockchain-Based Agriculture Assistance. Lect. Notes Electr. Eng. 2021, 700, 477.

[10] Prasath, J.S.; Jayakumar, S.; Karthikeyan, K. Real-time implementation for secure monitoring of wastewater treatment plants using internet of things. Int. J. Innov. Technol. Explor. Eng. 2019, 9, 2997–3002.

[11] Srisruthi, S.; Swarna, N.; Ros, G.M.S.; Elizabeth, E. Sustainable agriculture using eco-friendly and energy efficient sensor technology. In Proceedings of the 2016 IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), Bangalore, India, 20–21 May 2016; IEEE: Bangalore, India, 2016; pp. 1442–1446.

[12] Brodt, S.; Six, J.; Feenstra, G.; Ingels, C.; Campbell, D. Sustainable Agriculture. Nat. Educ. Knowl. 2011, 3, 1.

[13] Obaisi, A.I.; Adegbeye, M.J.; Elghandour, M.M.M.Y.; Barbabosa-Pliego, A.; Salem, A.Z.M. Natural Resource Management and Sustainable Agriculture. In Handbook of Climate Change Mitigation and Adaptation; Lackner, M., Sajjadi, B., Chen, W.Y., Eds.; Springer: Cham, Switzerland, 2022.

[14] Latake, P.T.; Pawar, P.; Ranveer, A.C. The Greenhouse Effect and Its Impacts on Environment. Int. J. Innov. Res. Creat. Technol. 2015, 1, 333–337.

[15] Reddy, T.; Dutta, M. Impact of Agricultural Inputs on Agricultural GDP in Indian Economy. Theor. Econ. Lett. 2018, 8, 1840–1853.

[16] World Agriculture: Towards 2015/2030: An FAO Perspective and Summary Report; FAO: Rome, Italy, 2002; Available online: www.fao.org/3/a-y4252e.pdf (accessed on 1 August 2022).

[17] Roser, M.; Ritchie, H.; Ortiz-Ospina, E. World Population Growth. 2013. Available online: https://ourworldindata.org/world-population-growth (accessed on 1 August 2022).

[18] Hernández-Ochoa, I.M.; Gaiser, T.; Kersebaum, K.C.; Webber, H.; Seidel, S.J.; Grahmann, K.; Ewert, F. Model-based design of crop diversification through new field arrangements in spatially heterogeneous landscapes. A review. Agron. Sustain. Dev. 2022, 42, 74.

[19] Navulur, S.; Sastry, A.S.C.S.; Giri Prasad, M.N. Agricultural Management through Wireless Sensors and Internet of Things. Int. J. Electr. Comput. Eng. 2017, 7, 3492–3499.

[20] Ayaz, M.; Ammad-uddin, M.; Baig, I.; Aggoune, E.M. Wireless Sensor's Civil Applications, Prototypes, and Future Integration Possibilities: A Review. IEEE Sens. J. 2018, 18, 4–30.

[21] H. Sharma, A. Haque, and Z. A. Jaffery, "Smart agriculture monitoring using energy harvesting Internet of Things (EH-IoT)," An International Scientific Journal, vol. 121, pp. 22–26, 2019.

[22] M. Suchithra, "Sensor data validation," International Journal of Pure and Applied Mathematics, vol. 119, no. 12, pp. 14327–14335, 2018.

[23] P. Joshi, "Wireless sensor network and monitoring of crop field," IOSR Journal of Electronics and Communication Engineering (IOSR-JECE), vol. 12, no. 1, pp. 23–28, 2017.