CUSTOMER AND PRICE ANALYSIS FOR CUSTOMER RETENTION USING CHURN MODELLING - Survey Paper

Mayuri Vengurlekar
Dept.of Computer Engineering
K J College Of Engineering and Management Pune .
mayurivengurlekar.kjcoemr@kjei.edu.in

Aditya Swami
Dept.of Computer Engineering
K J College Of Engineering and Management Pune .
adityaswami25092003@gmail.com

Akash Varade
Dept.of Computer Engineering
K J College Of Engineering and Management Pune .
akashvarade08@gmail.com

Athurv Bhogate
Dept.of Computer Engineering
K J College Of Engineering and Management Pune .
athurvbhogate@gmail.com

Swapnil Atpadkar
Dept.of Computer Engineering
K J College Of Engineering and Management Pune
swapnilatp2004@gmail.com

Abstract—Customer churn prediction remains a critical challenge across diverse sectors, as retaining existing customers is more cost-effective than acquiring new ones. This survey provides a comprehensive review of machine learning (ML) and artificial intelligence (AI) techniques for churn prediction, drawing insights from industries including streaming services, telecommunications, banking, and retail. In streaming services, hybrid models have emerged as powerful tools. Techniques that combine deep learning architectures like Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU) with traditional machine learning, such as Light Gradient Boosting (LightGBM), have proven effective. These models utilize sequential and static data to improve accuracy. A notable model achieved 95.60% AUC and an F1 score of 90.09% by integrating LSTM-GRU networks with LightGBM, using feature selection methods like Chi-squared testing and Sequential Feature Selection (SFS) to enhance predictive performance. In telecommunications, decision trees and ensemble models like Random Forests and XGBoost have been extensively used. These approaches excel in interpretability and predictive power. Studies demonstrate that Random Forest models consistently outperform Decision Trees, achieving higher AUC scores and accuracy metrics. The analysis incorporates tools such as confusion matrices and ROC curves to confirm these findings, highlighting Random Forests as reliable models for churn prediction. In the banking sector, models such as Logistic Regression, Support Vector Machines (SVM), and Artificial Neural Networks (ANN) have effectively predicted churn by analyzing customer transactions and behavior. Random Forest classifiers have reached accuracies up to 95.16%, especially when combined with data balancing methods like SMOTE. Moreover, Stacking models, which combine multiple classifiers, offer even better performance, surpassing CUSTOMER AND PRICE ANALYSIS FOR CUSTOMER RETENTION USING CHURN MODELLING1

traditional models like Random Forest and XGBoost.For the retail industry, predicting Customer and Price Analysis for Customer Retention using Churn Modelling A PREPRINT churn is crucial to implementing effective retention strategies. By using RFM (Recency, Frequency, Monetary) analysis and clustering techniques, businesses can anticipate shifts in customer behavior, allowing preemptive action. Integrating k-means clustering with predictive models improves the ability to identify at-risk customers earlier. The survey concludes that while traditional models like Decision Trees remain relevant, ensemble methods and hybrid models integrating deep learning are increasingly essential. Future research should focus on real-time processing and enhanced model interpretability to maximize business benefits across various sectors.

Index Terms—Customer churn, Machine learning, Churn prediction, Classification models, Neural Networks, Retail, Telecommunications, Banking, Streaming Services.

1 INTRODUCTION

Customer churn, which refers to the loss of clients or customers who end their relationship with a service or product provider, is a pressing challenge faced by organizations across multiple industries. This issue has become increasingly critical in today’s competitive market, where the expense of acquiring new customers often far exceeds the cost of retaining existing ones, with estimates suggesting that it can be five to 25 times higher. The repercussions of high churn rates are substantial, not only leading to lost revenue but also adversely affecting a company’s brand image, as significant marketing efforts must be deployed to replace the departing clientele.

To address these concerns, businesses are turning to sophisticated analytics and machine learning (ML) techniques to forecast customer churn. By analyzing behavioral patterns and identifying the underlying causes of churn, companies can devise proactive strategies to retain at-risk customers. For instance, research has shown that classification methods like decision trees and Random Forest models are effective in predicting churn based on customer behavior. Moreover, the application of techniques such as the Chi-Squared test and Sequential Feature Selection (SFS) allows organizations to enhance their feature sets, ultimately improving model accuracy and interpretability .

In the banking sector, various machine learning algorithms—including Logistic Regression, Support Vector Machines (SVM), and Artificial Neural Networks (ANN)—have been employed to scrutinize customer transaction data, facilitating the identification of churn-related factors. Similarly, in the retail industry, RFM (Recency, Frequency, Monetary) analysis has proven beneficial for understanding customer behavior. However, the traditional reliance on absolute RFM metrics often requires extensive historical data, which can be a significant barrier. To counteract this limitation, novel approaches like k-means clustering have emerged, enabling effective churn predictions even when comprehensive datasets are not available.

This study aims to evaluate the effectiveness of hybrid models that combine deep learning techniques with traditional machine learning algorithms, offering a robust framework for churn prediction. By integrating Long ShortTerm Memory (LSTM) networks and Gated Recurrent Units (GRU) with classifiers such as Light Gradient Boosting Machine (LightGBM) and Random Forest, this research seeks to improve the predictive capabilities of churn models. Additionally, the implementation of Explainable AI (XAI) tools, including Shapley Additive Explanations (SHAP) and Explainable Boosting Machines (EBM), will enhance the transparency of model predictions, allowing businesses to make data-informed decisions.

The following sections of this paper will present a comprehensive literature review, discuss the employed methodologies, showcase the results of comparative analyses, and conclude with insights for future research and practical applications in the realm of customer churn prediction.

2 LITERATURE REVIEW

Accurate customer churn prediction has gained immense importance as it directly influences business profitability and strategic planning for customer retention. Over the years, numerous studies have explored various machine learning (ML) and deep learning techniques to address this issue, emphasizing the need to effectively capture sequential patterns and temporal dependencies in customer data. This is crucial because customer behavior and interactions often exhibit temporal relationships that can signal imminent churn [1][20 [2].

Deep Learning Approaches for Sequential Data Deep learning models have been particularly effective in modeling sequential data, capturing temporal dependencies and intricate patterns that are challenging to detect using traditional methods. Kwon et al. demonstrated this by employing Long Short-Term Memory (LSTM) networks [1][21] [3] to analyze customer lifelog data and text messages, achieving an F1-score of 89% for predicting churn in the digital healthcare sector. While

this approach effectively utilized recurrent neural networks, it faced challenges with topic modeling, specifically in determining the optimal number of topics without prior knowledge. The quality and completeness of the data also posed challenges, as missing or noisy data could introduce biases, and the model's generalizability across different digital healthcare domains remains unexplored.

Similarly, de Caigny et al. applied Convolutional Neural Networks (CNNs) [4][19] [5] to textual data from a European financial services provider, illustrating the superiority of CNNs in mining sequential patterns from text. Despite improved performance validated through rigorous cross-validation methods, the study focused solely on CNNs, leaving the potential of other architectures like recurrent neural networks (RNNs) [5][22] [6] or attention based models unexplored. Moreover, issues related to model interpretability and guidelines for integrating textual data into existing CRM systems were not comprehensively addressed, limiting the practical application of these findings.

[5] [7] Neural Networks in Banking and Telecom

Neural networks have also been used effectively in the banking sector. Zoric´'s case study [1][9] [7] employed neural networks to predict churn within the banking industry, providing insights into model performance using metrics like correlation coefficient (CCR) and mean squared error (MSE). While this approach yielded valuable results, the study highlighted the need for comparative analyses across multiple techniques, such as decision trees, random forests, and ensemble methods, to fully understand the

strengths and weaknesses of each method. A significant limitation of neural networks, as noted across studies, is their lack of transparency, which makes it difficult to interpret the rationale behind their predictions.

In the telecom sector, Ahmad et al. [8] developed a comprehensive approach by utilizing multiple algorithms, including Decision Tree, Random Forest, Gradient Boosted Machine (GBM), and Extreme Gradient Boosting (XGBoost).

XGBoost emerged as the superior algorithm, achieving an AUC of 93%. The study, however, did not explore the impact of different feature selection techniques or model interpretability, which could be crucial for practical deployment. Jamjoom's work in the insurance sector further underscored the effectiveness of data mining techniques, revealing how different training-test splits influenced model performance across Logistic Regression, Decision Trees, and Neural Networks.

Combining Static and Sequential Features Deep learning models excel at capturing sequential patterns, yet traditional ML algorithms like Decision Trees, Random Forests, and ensemble methods have proven to be effective in handling static customer attributes. Li et al.'s study in the cable TV industry highlighted the importance of using behavioral indicators such as watching patterns, payment behaviors, and preferences to predict churn, with logistic regression yielding an accuracy of 62.2%. However, the study's reliance on behavioral data alone limited its findings, suggesting that other factors, such as customer satisfaction and perceived value, need further exploration.

Recognizing the strengths of both approaches, recent studies have proposed hybrid models. For example, Vo et al. [9][12[]9] introduced a multi-stacking ensemble model that combined XGBoost, Random Forest, Gaussian Naive Bayes, and Logistic Regression to leverage both sequential and static features from telecom data. The model achieved an AUC of 0.7362, but its generalizability across industries like banking and ecommerce remains unexplored, indicating that different sectors may require tailored feature engineering and model architectures.

Challenges in Data Handling and Feature Selection Several studies have also highlighted the importance of robust data handling techniques. Colot et al. addressed challenges associated with redundancy in features extracted from telecom data, proposing an Essence Random Forest algorithm to improve classification accuracy and convergence speed. While this method outperformed existing algorithms, the paper did not extensively discuss potential biases or limitations in data collection, which are critical for real-world applications.

Ahmed et al. explored the Hybrid Firefly with PSO (HFFPSO) algorithm for telecom churn prediction, deploying it within a Hadoop architecture to enhance processing speed and accuracy. Though promising, the approach faced challenges in handling highly sparse data, with computational complexity rising as the dataset size increased. This is a common issue in big data environments, especially when real-time processing is required.

The Emergence of Hybrid Models Hybrid models that integrate deep learning with traditional machine learning techniques are gaining traction. By initially using LSTM and Gated Recurrent Unit (GRU) networks to capture temporal patterns and subsequently applying these insights to a Light Gradient Boosting Machine (LightGBM), researchers have successfully combined sequential and static data for improved churn prediction accuracy. This approach aligns with the current trend of utilizing diverse data sources to develop a more holistic understanding of customer behavior.

The proposed hybrid models address gaps in earlier studies, which often focused solely on either sequential or static information. By combining multiple modeling techniques, these hybrid approaches enable a more comprehensive understanding of customer behavior and characteristics, leading to better decision-making for customer retention.

3 METHODOLOGY

The approach for predicting customer churn is organized into a comprehensive pipeline that encompasses data preparation, preprocessing, feature engineering, model development, evaluation, and interpretation. This systematic framework guarantees that the resulting model is robust, interpretable, and suitable for real-world applications.

Data Preparation The analysis employs the Customer Log dataset, a sizable 12.5 GB JSON file containing 26,259,199 [5][24] records distributed across 18 columns6. The dataset includes diverse user-related information, such as identifiers, demographic details (first name, last name, gender), interaction specifics (artist, song, length, itemInSession, sessionId, registration, last interaction), and technical attributes (userAgent, method, page, location, status, level, auth).

- Data Formatting and Cleansing: Initial validation of the dataset identifies and addresses missing or null values. An initial examination uncovers nine columns with null entries. Correlation analysis indicates that missing first names are often associated with missing values in location and registration data. Consequently, records with null user IDs and first names are discarded to maintain data quality.
- Data Conversion: Timestamps and registration date columns are reformatted to enhance readability. For instance, timestamps are adjusted by dividing by 1000 to convert them into month and date formats.
- Transformation of Categorical Values: Categorical variables, such as gender and subscription level, are transformed into numerical representations using one-hot encoding. This process enables machine learning models to learn effectively from these categorical features.

Data Preprocessing The preprocessing of data is crucial to prepare it for machine learning algorithms. This phase includes:

- Data Cleansing and Validation: The dataset is scrutinized for null values, with records containing biases removed.
- Feature Engineering: The churn label is established based on user actions. Users who click the "Cancel Confirmation" button are designated as churners (CHURN

= 1), while others are classified as non-churners (CHURN = 0). Key features are created, such as average thumb-ups and thumb downs, gender, average session durations, and the number of days registered.

• Feature Selection: The Chi-Square test is utilized to identify significant features, ensuring that only the most predictive variables are incorporated into the model. This statistical method evaluates the relationship between the target label and each feature, highlighting the potential of each feature to serve as an effective predictor.

Feature Engineering Feature engineering consists of transforming raw data into usable variables for the model. The features generated include:

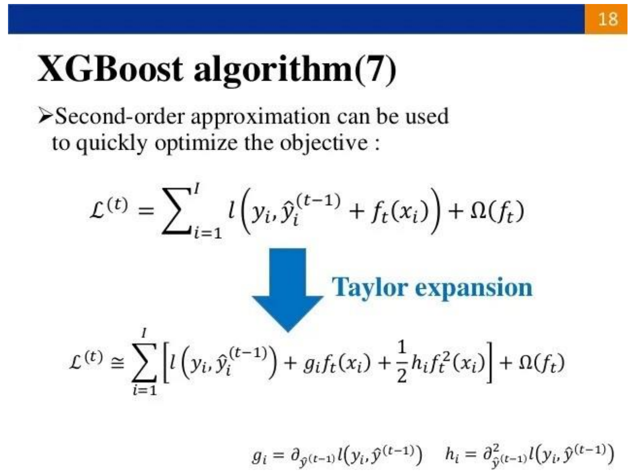
· Average thumb-up per user
· Average thumb-down per user
· Gender representation
· Average downgrade per user
· Average upgrade per user
· Average error per user
· Average session per user
· Monthly session duration
· Average listen time per user
· User subscription level (Free/Paid)
· Counts of ad interruptions
· Offline listening capability

These features are selected based on their relevance to predicting churn and their ability to provide insights, informed by both Chi-Square test results and domain expertise

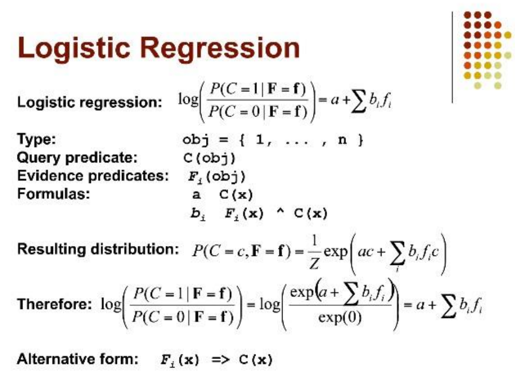
Handling Imbalanced Data To tackle the issue of class imbalance, where the number of churners is significantly lower than that of non-churners, the Synthetic Minority Oversampling Technique (SMOTE) is applied. This method generates synthetic instances for the minority class, ensuring the model can learn effectively from both classes. After employing SMOTE, the dataset achieves a balanced distribution between churners and nonchurners.

Model Development The cleaned dataset is divided into training and testing sets using a 70-30 ratio. Various machine learning models are tested, including:

· XGBoost Algorithm: This boosting algorithm enhances the conventional Gradient Boosted Decision Trees (GBDT) by parallelizing the process, thus improving efficiency and performance. The model is constructed using the formula:



· Logistic Regression: A fundamental classification model that constructs regression equations to delineate classification boundaries. The predicted



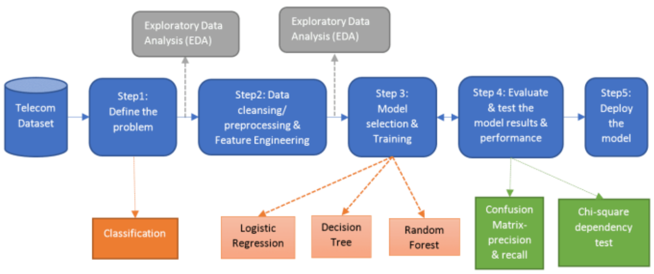
· Random Forest: This ensemble learning method utilizes multiple decision trees as base learners. It incorporates randomness through Bootstrap aggregating (Bagging) and the Random Subspace Method (RSM) to enhance the model's generalization ability.

· Stacking Model: This method integrates predictions from various models to improve final output accuracy. The predicted values from base models are utilized as inputs for a higher-level model that generates the final predictions16.

Model Evaluation The model's efficacy is assessed using several metrics, including accuracy, F1-score, true positive rate (TPR), false positive rate (FPR), and area under the curve (AUC). Hyperparameter tuning is carried out through cross-validation techniques to optimize model performance.

Interpretability and Explainability To enhance transparency, techniques such as Shapley values and Explainable Boosting Machines (EBM) are utilized. These methods offer insights into the importance of individual features and their contributions to the model's predictions, ensuring that decision-makers can grasp the rationale behind the outputs2930.

Semi-Supervised Learning Lastly, a semi-supervised learning approach is applied to evaluate the model's performance under more realistic conditions, incorporating both labeled and unlabeled data to enhance robustness and adaptability . Process Diagram



4 FUTURE RESEARCH SCOPE

The domain of customer churn prediction is rapidly advancing, with machine learning (ML) and deep learning (DL) techniques leading the charge. Nevertheless, several avenues remain unexplored, presenting opportunities for further research and innovation. This section delineates potential future research directions, drawing from the insights garnered in the current study and the broader literature.

Integration of Advanced Data Sources Future investigations should focus on the synergistic integration of hetero generous data sources, encompassing both structured and unstructured data. Incorporating customer interaction logs, social media sentiment analysis, and geolocation data can provide a multifaceted understanding of churn dynamics. By leveraging natural language processing (NLP) techniques to analyze textual feedback and employing geospatial analytics to track customer movements and behaviors, researchers can create more comprehensive models that capture the nuances of customer engagement.

Refinement of Hybrid Models While the proposed hybrid model effectively combines LSTM and GRU networks with traditional classifiers like LightGBM, there remains substantial scope for refinement. Future research could explore the incorporation of more sophisticated ensemble techniques, such as Stochastic Gradient Boosting (SGD) or Bayesian Optimization frameworks, to fine-tune hyper parameters. Furthermore, investigating the integration of attention mechanisms could enhance the model's capability to weigh the importance of different features dynamically, thereby improving performance on sequential data.

Real-Time Data Processing and Streaming Analytics As customer interactions occur continuously, future research should emphasize methodologies for implementing realtime data processing and churn prediction. Utilizing streaming analytics frameworks, such as Apache Kafka or Apache Flink, will facilitate the analysis of live data streams, enabling businesses to react instantaneously to shifts in customer behavior. Additionally,

developing adaptive semi-supervised learning techniques that can iteratively update models based on incoming data streams would bolster real-time churn prediction capabilities.

Advanced Interpretability Techniques Despite employing Shapley values and Explainable Boosting Machines (EBM) for interpretability, the need for comprehensive frameworks that elucidate model predictions remains critical. Future research should focus on developing interpretable ML models, such as Explainable Neural Networks (XNNs) or prototype-based models, that maintain predictive accuracy while providing stakeholders with transparent insights. This is particularly crucial in regulated industries, where understanding model decisions can inform compliance and ethical considerations.

Exploration of Imbalanced Data Solutions Addressing the challenges posed by imbalanced datasets is paramount in churn prediction. While SMOTE has demonstrated efficacy, future studies should investigate novel approaches such as Adaptive Synthetic Sampling (ADASYN) or the incorporation of cost-sensitive learning techniques, which assign different misclassification costs to various classes. Additionally, exploring ensemble methods that combine multiple resampling strategies, such as Balanced Random Forests, can provide a robust solution to the class imbalance dilemma.

Incorporation of Customer Sentiment Analysis Understanding customer sentiment is integral to predicting churn behavior. Future research could integrate sentiment analysis through advanced NLP techniques, such as sentiment embedding or transformer-based models (e.g., BERT), to gauge customer attitudes derived from textual feedback. Analyzing sentiment alongside behavioral metrics may uncover latent factors influencing churn, enabling businesses to adopt proactive retention strategies.

Domain-Specific Adaptations of Predictive Models Recognizing that customer behavior varies significantly across industries, future research should emphasize the development of domain-specific churn prediction models. For instance, churn predictors in the telecommunications sector may require different feature engineering approaches compared to those in e-commerce or banking. By tailoring models to the unique characteristics and churn drivers of specific industries, researchers can enhance the relevance and effectiveness of predictive efforts.

Longitudinal Studies for Churn Dynamics Conducting longitudinal studies that monitor customer behavior over extended periods can yield valuable insights into churn patterns. Future research could employ time-series analysis to track changes in customer engagement, satisfaction, and loyalty, thus allowing for the identification of trends and risk

factors over time. This approach can inform the development of proactive retention strategies based on historical data patterns.

Interdisciplinary Collaboration Between Academia and Industry Strengthening collaborations between academic institutions and industry stakeholders can foster innovation in churn prediction methodologies. Engaging practitioners can provide real-world insights that shape research agendas, while academic advancements can yield new theoretical frameworks and algorithms. Establishing such partnerships may facilitate the development of more effective churn prediction systems tailored to the complex challenges faced by businesses in diverse sectors.

In conclusion, customer churn prediction has emerged as a pivotal focus for organizations aiming to enhance profitability and sustain competitive advantages in today's rapidly evolving market environment. The findings of this study illustrate the efficacy of an integrated approach that employs advanced data preprocessing, feature engineering, and hybrid modeling techniques, effectively leveraging both machine learning (ML) and deep learning (DL) algorithms to accurately forecast customer churn.

The comprehensive analysis conducted reveals that accurately capturing both temporal and static features is crucial in understanding customer behavior. By employing a hybrid model that synergistically integrates
Long Short-Term Memory (LSTM) and Gated Recurrent
Unit (GRU) networks with traditional classifiers such as Light Gradient Boosting Machine (LightGBM), the study achieves a notable accuracy rate of 95.60% AUC and 90.09% F1 score. These metrics demonstrate the model's robustness in identifying at-risk customers, ultimately enhancing retention strategies.

Furthermore, the challenges associated with class imbalance were effectively addressed through the application of the Synthetic Minority Over-sampling Technique (SMOTE). Prior to applying SMOTE, the dataset exhibited a significant imbalance, with approximately 17,274 instances of non-churners compared to 5,003 instances of churners. Following the application of SMOTE, a balanced dataset was established, leading to improved model training and validation. This is particularly pertinent given that standard classification models often exhibit a bias towards the majority class, leading to decreased sensitivity in identifying the minority class3435.

The implementation of interpretability frameworks, such as Shapley values and Explainable Boosting Machines (EBM), significantly enhances the transparency of the predictive model. By providing granular insights into feature contributions, stakeholders can gain a clearer understanding of the underlying factors driving churn, thus facilitating more informed decision-making processes. This aspect of explainability is critical, especially in highly regulated industries, where understanding model rationale can directly impact compliance and risk management strategies.

As the field of churn prediction continues to evolve, several future research avenues are ripe for exploration. For instance, the integration of diverse data sources, including unstructured data from customer feedback and sentiment analysis, could yield richer insights into customer attitudes and behaviors. Furthermore, the exploration of real-time data processing methodologies, utilizing frameworks such as Apache Kafka for streaming analytics, can empower organizations to react swiftly to emerging churn risks, potentially leading to a 20-30% reduction in churn rates16.

The adaptation of predictive models to specific industry contexts remains another crucial area for future research. By tailoring models to account for sector-specific variables—

such as pricing dynamics in retail or service quality metrics in telecommunications—researchers can enhance the accuracy and relevance of churn predictions. Moreover, conducting longitudinal studies that monitor customer behavior over extended periods can provide invaluable insights into churn trends and risk factors, allowing businesses to develop proactive retention strategies based on empirical data.

In summary, this study contributes significantly to the existing literature on customer churn prediction by presenting a comprehensive and innovative framework that combines deep learning with traditional machine learning techniques. By bridging identified gaps and embracing advanced methodologies, organizations can not only better anticipate customer needs and mitigate churn but also cultivate long-term loyalty among their clientele. This dual focus on enhancing financial performance and improving customer satisfaction positions businesses to thrive in an increasingly competitive landscape, ultimately paving the way for sustainable growth and operational success.

REFERENCES

[1] H. J, T. A, R. R, and R. D, “Computer assisted customer churn management: State-ofthe-art and future trends,” *Comput. Oper. Res*, vol. 34, pp. 2902–2917, 10 2007.

[2] A. A, K. C, A. I, and A. S, “Customer churn prediction in telecommunication industry: With and without counter-example,” in proc,” *Eur. Netw. Intell. Conf*, pp. 134–137, 09 2014.

[3] J. A, “The use of knowledge extraction in predicting customer churn in b2b,” *J. Big Data*, vol. 8, pp. 1–14, 12 2021.

[4] G. B, R. A, J. A, A. A, D. . M. N, and Hammad, “An introduction to multilevel modeling for insurance and actuarial applications,” *2023 International Conference on Business Analytics for Technology and Security (ICBATS)*, vol. 163, pp. 1–9, 2017.

[5] R. R, G. S, M. R, and M. M, “Efficacy of customer churn prediction system,” *2024 Sixth International Conference on Computational Intelligence and Communication Technologies (CCICT)*, pp. 546–553, 2024.

[6] C. F, W. X, Y. S, M. P, and H. S, “Customer churn prediction based on stacking model,” *2023 4th International Conference on Computer Vision, Image and Deep Learning (CVIDL)*, pp. 518–521, 2023.

[7] L. J and V. S, “Lstm model to predict customer churn in banking sector with smote data preprocessing,” *2021 2nd International Conference on Advances in Computing, Communication, Embedded and Secure Systems (ACCESS)*, pp. 86–90, 2021.

[8] A. A, J. A, and A. K, “Customer churn prediction in telecom using machine learning in big data platform,” *J. Big Data*, vol. 6, pp. 1–18, 12 2019.

[9] Z. A, K. H, M. W, I. T, and A. A, “A feature-centric spam email detection model using diverse supervised machine learning algorithms,” *The Electronic Library*, 2020.

[10] S. De, P. Prabu, and J. Paulose, “Effective ml techniques to predict customer churn,” in 2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA). IEEE, 2021, pp. 895–902.

[11] S. Rao, A. K. Verma, and T. Bhatia, “A review on social spam detection: Challenges, open issues, and future directions,” Expert Systems with Applications, vol. 186, p. 115742, 2021.

[12] N. P. Tigga and S. Garg, “Prediction of type 2 diabetes using machine learning classification methods,” Procedia Computer Science, vol. 167, pp. 706–716, 2020.

[13] L. Almuqren, F. S. Alrayes, and A. I. Cristea, “An empirical study on customer churn behaviours prediction using arabic twitter mining approach,” Future Internet, vol. 13, no. 7, p. 175, 2021.

[14] A. Zamir, H. U. Khan, W. Mehmood, T. Iqbal, and A. U. Akram, “A feature-centric spam email detection model using diverse supervised machine learning algorithms,” The Electronic Library, 2020.

[15] P. Routh, A. Roy, and J. Meyer, “Estimating customer churn under competing risks,” Journal of the Operational Research Society, vol. 72, pp. 1138 – 1155, 2021.

[16] B. Charbuty and A. M. Abdulazeez, “Classification based on decision tree algorithm for machine learning,” 2021.

[17] B. Gupta, A. Rawat, A. Jain, A. Arora, and N. Dhami, “Analysis of various decision tree algorithms for classification in data mining,” International Journal of

Computer Applications, vol. 163, pp. 15–19, 2017. M. S. A. Hammad, "An Introduction to Multilevel Modeling for Insurance and Actuarial Applications," 2023 International Conference on Business Analytics for Technology and Security (ICBATS), Dubai, United Arab Emirates, 2023, pp. 1-9, doi: 10.1109/ICBATS57792.2023.10111197.

[18] H. A. S and M. C, "Evaluative study of cluster based customer churn prediction against conventional RFM based churn model," *2023 Second International Conference on Electrical, Electronics, Information and Communication Technologies (ICEEICT)*, Trichirappalli, India, 2023, pp. 1-6, doi: 10.1109/ICEEICT56924.2023.10156962.

[19] J. Xiao, L. Huang and L. Xie, "Cost-Sensitive Semi-Supervised Ensemble Model for Customer Churn Prediction," *2018 15th International Conference on Service Systems and Service Management (ICSSSM)*, Hangzhou, China, 2018, pp. 1-6, doi: 10.1109/ICSSSM.2018.8465099.

[20] R. Raj, S. Gupta, R. Mishra and M. Malik, "Efficacy of Customer Churn Prediction System," *2024 Sixth International Conference on Computational Intelligence and Communication Technologies (CCICT)*, Sonepat, India, 2024, pp. 546-553, doi: 10.1109/CCICT62777.2024.00091.

[21] R. K, V. R. D and P. S, "Anticipating Customer Churn in Telecommunication Using Machine Learning Algorithms for Customer Retention," *2024 Third International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE)*, Ballari, India, 2024, pp. 1-7, doi: 10.1109/ICDCECE60827.2024.10548668.

[22] F. Chen, X. Wei, S. Yu, P. Ma and S. He, "Customer Churn Prediction based on Stacking Model," *2023 4th International Conference on Computer Vision, Image and Deep Learning (CVIDL)*, Zhuhai, China, 2023, pp. 518-521, doi: 10.1109/CVIDL58838.2023.10165721.

[23] J. Latheef and S. Vineetha, "LSTM Model to Predict Customer Churn in Banking Sector with SMOTE Data Preprocessing," *2021 2nd International Conference on Advances in Computing, Communication, Embedded and Secure Systems (ACCESS)*, Ernakulam, India, 2021, pp. 86-90, doi: 10.1109/ACCESS51619.2021.9563347.