JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

CUSTOMER RETENTION AND CHURN PREDICTION QUANTITATIVE AND QUALITATIVE APPROACH IN BANKING INDUSTRY

[1]Ms.S. Kavitha, [2]Alwin Sanjai Z, [3] Dhinesh J, [4] Duvvuri Abhishek, [5] Dominic P
[1]Assistant professor, [2]Student, [3]Student,[4] Student, [5] Student
[1]Computer science engineering,
[1]Vel Tech High Tech Dr Rangarajan Dr Sakunthala Engineering College, Avadi, India

Abstract :

Large amounts of data have been generated as a result of the advancement of technology in the modern era. The 2.5 quintillions of data produced daily by people with Internet connections in 2023 are proof of this. By 2024, there will likely be 5.3 billion Internet users worldwide. This means that sophisticated and effectivemodels, tools, or techniques are required to investigate, evaluate, and extract valuable hidden information frommassive amounts of data. Machine learning methods including clustering, decision trees, and logistic regressionhave become more important inrecent years, particularly in churn prediction. The technique of estimating thepercentage of customers who avoid using or might cease subscribing to a product or service provided by anorganization or company is known as customer churn prediction. Though various prediction models have beenproposed, most research attention has been given to measuring the efficiency of prediction models, rather thanidentifying its application for sustainable economic development. In this paper, we investigate the determiningfactor for customer attrition in the banking sector using Power BI, Tableau. Dataset from world bank spain waspreprocessed with four key client factors were utilized. The LRRFSGB calculation accessible within the ControlBi,Tableau program was utilized for preparing and testing. The comes about appear that customer account adjustcould be a key deciding variable for churning. Moreover, the comes about appear that churning happens less inmale than female clients. This work will providebanks with valuable information on building successful clientmaintenance procedures. Building an viable and precise client churn forecast demonstrate is an critical inquire about issue for both scholastics and professionals.

IndexTerms **-.Customer churn, KNN, LRRFSGB algorithms, Power BI , Tableau**

1.INTRODUCTION

In todays rapidly evolving technological landscape, the accuracy of tools andalgorithms plays a crucial role in determining the success of various projects across diverse domains. As such, yourproject focused banking industries on systematically evaluating the accuracy of these tools and algorithms toensure their reliability and effectiveness in real-world applications .Many industries build a model like a churn as acommon application for data mining technique. Mobile telephone organizations present across the globe arealmost on the verge of building their own churn model[].The remainder of the study is organized as follows: In thefollowing subsection, we present literature review and overview of the DTI. Section 2 describes four datasets andthe methodologies used in this study.[1]By employing a structured approach that encompasses defining clearobjectives, selecting appropriate tools and algorithms, gathering and preprocessing relevant data, and rigorouslyevaluating performance metrics, we aim to provide valuable insights into the strengths, weaknesses, and overallperformance of the technologies under scrutiny. Ensemble learning involves using several individual classifiers andcombining their predictions, which may result in better performance than a single classifier [4].Through thisendeavor, we seek to empower decision-makers with actionable intelligence and pave the way for advancementsin technology-driven solutions. The managing an account industry faces the chalenge of holding clients. Clientsmay switch to another bank for different reasons, such as superior monetary administrations due to moo rates,bank department areas, quality of computerized instruments, and low-interest rates, among others (Kaur & Kaur,

2020). The web and the blast of social media have had a colossal affect on the retail keeping money industry overprogressively digitized, and this drift shows up to be relentless. The current ponder is significant since it could be afollow-up to the company's current churn-related work ventures. The bank is as of now examining the clientlifetime esteem, with churn expectation being an critical component of this ponder since it contributes tocustomer characterization. Client Lifetime Esteem (CLV) could be a vital pointer that companies ought to screenand control. It helps in their understanding of the budgetary misfortunes caused from losing clients. Furthermore,it helps them in calculating the cost of drawing in modern clients and the potential benefit from each one. Theobjective of this investigate was to help within the improvement of unused items and administrations that would include esteem to the company's show clients and extend its relationship with them. The completion of such aponder illustrates that churn could be a subject that causes broad stress inside the firm. Churn will proceed toexist, and performing client administration is superior than obtaining modern customersto guarantee longtermcommerce development and productivity Besides, in any case of industry, checking churn ought to be a consistentstress for any firm nowadays, since competition is extreme, and the advanced era has made it less demanding forclients to require their trade somewhere else in case they want. As a result, it is to any company's best advantageto keep up a observe on its customers' behavior to anticipate any despondency that might lead to churning. Suchsteps may be significant in coming to out to those clients and, perhaps, saving the bank's relationship. The moneyrelated rewards of such activities are self-evident, as potential churners select to keep their accounts with thebank. Moreover, accurately foreseeing the churners and persuading them to remain at the company canincrement income, indeed in the event that a churn expectation show produces a certain number of falsepositives

3. METHODOLOGY:

The methodology for customer churn analysis in banking encompasses several key steps.Firstly, it's crucial to establish a clear definition of churn, whether it's account closures,decreased activity, or fund transfers to other institutions. Data collection involves gathering comprehensive customer data, includingmn transaction history and demographics, followed by preprocessing to clean and prepare the data for analysis.Feature selection and engineering help identify relevant prdictors of churn,

while exploratory data analysisreveals patterns and insights. Model selection involves choosing appropriate, mahine learning algorithmsbalancing accuracy with interpretability. After training and evaluation, the model predicts churn probabilities for current customers. Implementation involves integrating predictions into business taking measures to retain at-risk customers. Continuous monitoring and iteration refine the methodology, ensuringongoing effectiveness and ethical considerations such as fairness and privacy protection are addressed throughout process.

3.1 TABLEAU :

Tableau is a leading data visualization and analytics software known for its intuitive interface and powerfulcapabilities. It enables users to connect to various data sources, including databases, spreadsheets, and cloudservices, allowing them to work with diverse datasets without complex integration processes. With Tableau, usescan create visually appealing charts, graphs, maps, and dashboards using a drag-and-drop interface. The softwareoffers extensive interactivity features, enabling users to explore data dynamically by filtering, drilling down, andhighlighting specific data points. Tableau also supports advanced analytics, allowing users to perform calculations,apply statistical functions, and conduct predictive analysis directly within the platform. Additionally, Tableauprovides collaboration and sharing capabilities through its Server and Online platforms, facilitating teamwork andknowledge sharing within organizations.

3.2 POWER BI :

Power BI is a robust business analytics tool developed by Microsoft, offering a suite of features for datavisualization, exploration, and sharing. It allows users to connect to a wide range of data sources, includingdatabases, cloud services, and online platforms, facilitating seamless integration of diverse datasets. With PowerBI, users can create interactive reports, dashboards, and visualizations using a user-friendly interface and drag-and-drop functionality. The software provides powerful data modeling capabilities, enabling users to transformand manipulate data to derive meaningful insights. Power BI's rich visualization options include various charttypes, maps, and custom visuals, allowing users to present their data in compelling and informative ways.Moreover, Power BI offers advanced analytics features such as predictive modeling and machine learningintegration, empowering users to uncover trends, patterns, and correlations within their data. With its cloud-basedarchitecture, Power BI enables real-time collaboration and sharing of insights across teams and organizations.

3.3 MY SQL WORK BENCH : MySQL Workbench is a powerful integrated development environment (IDE) for MySQLdatabase management and development. Developed by Oracle Corporation, MySQL Workbench provides asuite of tools designed to streamline database design, modeling, querying, administration, and migration

tasks. One of its key features is the visual database design tool, which allows users to create, edit, andvisualize database schemas using an intuitive graphical interface. This tool simplifies the process of designingcomplex database structures by enabling users to visually create tables, define relationships, and setconstraints.

3.4 KNN ALGORITHM :

The K-Nearest Neighbors (KNN) algorithm is a popular supervised machine learningalgorithm used for classification and regression tasks. It is based on the principle of similarity, where datapoints are classified or predicted based on the majority vote or average of their nearest neighbors in thefeature space.In KNN classification, the algorithm assigns a class label to a new data point by consideringthe class labels of its K nearest neighbors. The choice of K, the number of neighbors, is a crucialparameter in KNN, as it affects the algorithm's performance and generalization ability. A smaller value ofK may lead to overfitting, while a larger value may result in underfitting.The k-NN algorithmDenote the set of the k nearest neighbors of x as Sx. Formally Sx is defined as $Sx \subseteq D$ s.t. $|Sx|=k$and $\forall(x',y')\in D\setminus Sx$, $dist(x,x')\geq max(x'',y'')\in Sxdist(x,x'')$, (i.e. every point in D but not in Sx is at least as far away from xas the furthest point in Sx)

3.5 RANDOM FOREST ALGORITHM :

Random Forest Algorithm widespread popularity stems from its user-friendly nature and adaptability,enabling it to tackle both classification and regression problems effectively. The algorithm's strength lies in itsability to handle complex datasets and mitigate overfitting, making it a valuable tool for various predictive

tasks in machine learning.One of the most important features of the Random Forest Algorithm is that it canhandle the data set containing continuous variables, as in the case of regression, and categorical variables, asin the case of classificaion. It performs better for classification and regression tasks. In this tutorial, we willunderstand the working of random forest and implement random forest on a classification task.

3.5 LOGISTIC REGRESSION ALGORITHM :

Logistic regression is a supervised machine learning algorithm that accomplishes binary classification tasks bypredicting the probability of an outcome, event, or observation. The model delivers a binary or dichotomousoutcome limited to two possible outcomes: yes/no, 0/1, or true/false.Logical regression analyzes the relationshipbetween one or more independent variables and classifies data into discrete classes. It is extensively used inpredictive modeling, where the model estimates the mathematical probability of whether an instance belongs to aspecific category or not.

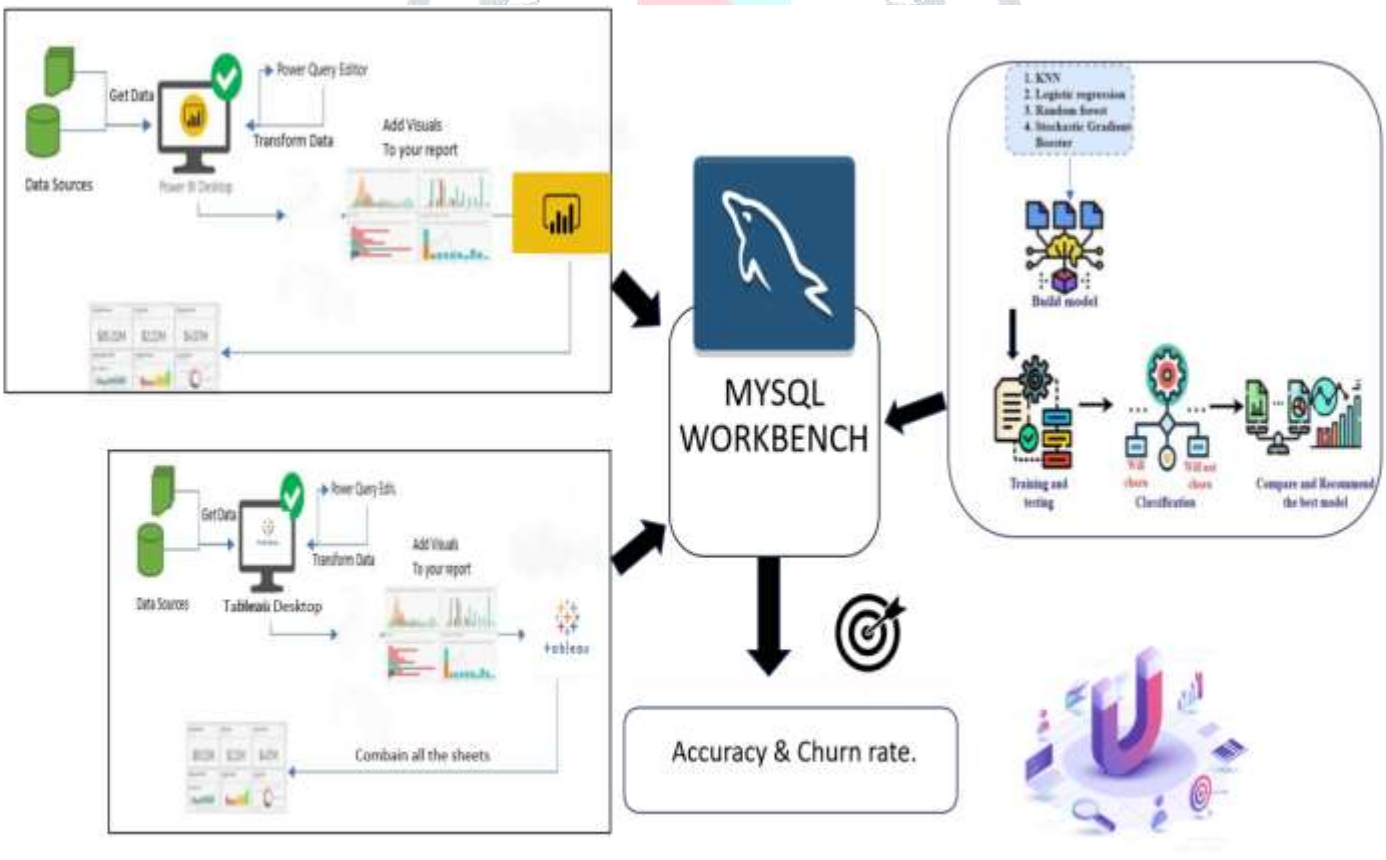

Figure:1 Diagram for constructing and finding the accuracy of tools and algorithms

ANALYSIS:

*Customer churn analysis for banking using Tableau and Power BI involves leveraging the capabilitieof these powerful visualization tools to gain actionable insights from data. These platforms offer intuitive interfacesand robust visualization features, enabling banks to conduct in-depth analyses of customer behavior and churnpatterns. With Tableau and Power BI, banks can create interactive dashboards and reports that highlight key metricsrelated to churn, such as customer demographics, transaction history, and engagement levels. By visualizing churntrends over time and identifying correlations between different factors, banks can uncover underlying reasons for

customer attrition and make informed decisions to address them. Furthermore, these tools facilitate dynamicexploration of data, allowing users to drill down into specific segments or anomalies that may require attention. Byharnessing the analytical power of Tableau and Power BI, banks can enhance their understanding of customer churndynamics and develop targeted retention strategies to mitigate churn effectively.

Figure 2: Dashboard in tableau

There are two numbers indicating the presence of customers in one country, with a breakdown by gender: 1,193 female customers and 1,316 male customers.A bar chart showing the age distribution of 2,509 customers. The x-axis represents age groups (bins), while the y-axis likely represents the number of customers in each age group.Top-right corner: A bar chart showing the age distribution of 1,248 active customers, which is 49.7% of the total customer base. There's also a number indicating that 814 customers have exited (no longer active customers). A bubble chart representing the number of customers and the number of exited customers in three countries: France, Germany, and Spain. The size of the bubbles likely corresponds to the number of customers, with larger bubbles indicating a higher number of customers.A bar chart showing the breakdown of customers by gender and whether they have exited (yes or no). The x-axis represents the gender (female or male), and the y-axis represents the number of customers. There are two sets of bars for each gender, one for customers who have not exited (no) and one for those who have (yes).Bottom-right corner: A legend indicating the measures used in the charts, such as age, balance, credit score, salary, product, and tenure.Overall, the dashboard provides insights into the customer base of a company, including their age distribution, gender, active status, credit card ownership, and geographical distribution. It's a tool for analyzing customer data to inform business decision.

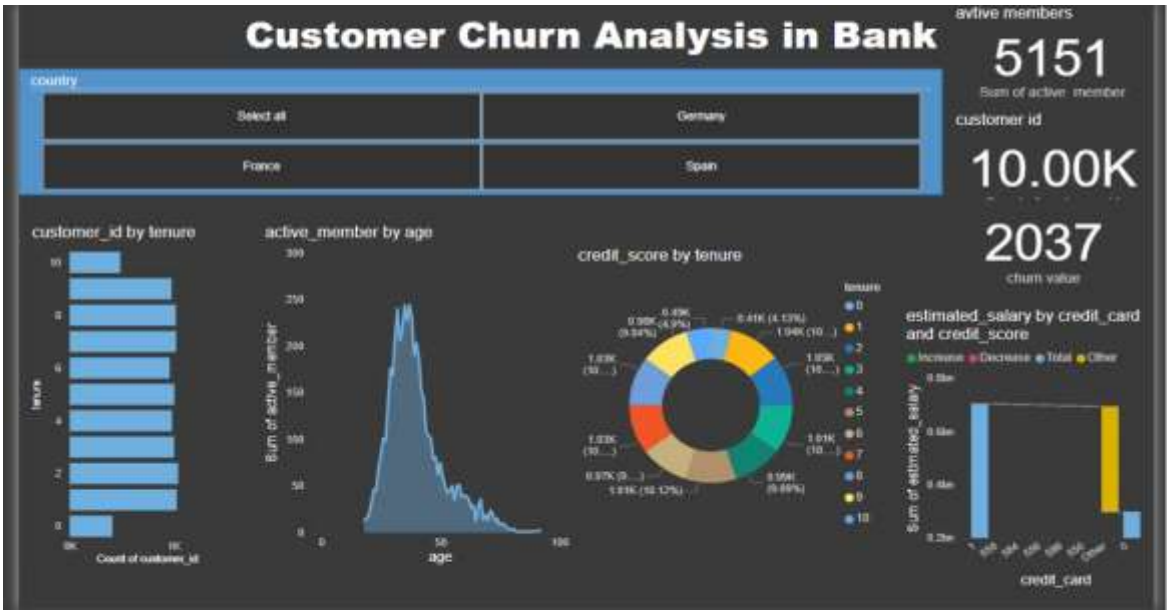

Figure 3: Dashboard in powerbi

POWER BI:

The title at the top indicates that the dashboard is focused on analyzing customer churn within a banking context.
On the top left, there is a filter for the country, allowing the user to select data for France, Germany, Spain, or all countries combined.On the bottom left, there is a horizontal bar chart showing the count of customers by tenure (how long they have been with the bank). The x-axis represents the count of customers, and the y-axis represents different tenure lengths. In the center, there is a line chart showing the number of active members by age. The x-axis represents the age of customers, and the y-axis represents the count of active members.On the bottom center, there is a donut chart showing the distribution of credit scores by tenure. Each segment of the donut represents a different tenure group, and the size of each segment indicates the proportion of customers within that tenure group. The colors likely correspond to different ranges of credit scores.
On the top right, there is a numerical value labeled "churn value," which is 2037. This number likely represents the total value (possibly in monetary terms) associated with customers who have churned.Also on the top right, there is a numerical value indicating the sum of active members, which is 5151.Below the sum of active members, there is a numerical value of 10.00K, which might represent the total number of customers or a specific subset of customers.On the bottom right, there is a bar chart showing the estimated salary distribution by credit card ownership and credit score. The x-axis represents whether customers have a credit card, and the y-axis represents the estimated salary. The bars are segmented by credit score ranges, indicated by different colors.
A legend on the bottom right explains the color coding used in the bar chart for different ranges of credit scores.Overall, this dashboard provides a visual representation of various metrics that are important for understanding customer behavior, particularly in relation to churn. It allows bank analysts to quickly assess the demographics of their customer base, including age, tenure, credit score, and salary, and how these factors might relate to customer retention

For effective customer churn analysis in banking using Tableau and Power BI, several recommendations can optimize the process and outcomes. Firstly, it's essential to design interactive dashboards and reports that provide a holistic view of churn metrics, including customer demographics, transaction history, and engagement levels. Incorporating dynamic filters and drill-down functionalities allows users to explore data at various levels of granularity, facilitating deeper insights into churn patterns. Additionally, leveraging advanced visualization techniques such as heatmaps, treemaps, and scatter plots can help identify correlations and trends that may not be apparent through traditional analysis methods.

*Furthermore, integrating external data sources such as market trends, competitor analysis, and customer feedback can enrich the analysis and provide context for churn behavior. Utilizing predictive analytics models within Tableau and Power BI enables banks to forecast future churn probabilities and prioritize retention efforts accordingly. Moreover, implementing real-time monitoring capabilities allows banks to promptly identify and respond to emerging churn risks, enhancing proactive retention strategies.

CONCLUSION:

*In conclusion, employing Tableau and Power BI for customer churn analysis in banking yields significant advantages in understanding and mitigating churn risks. Through interactive dashboards and reports, these platforms provide comprehensive insights into customer behavior, enabling banks to identify churn patterns, trends, and contributing factors effectively. By leveraging advanced visualization techniques and predictive analytics models, banks can forecast future churn probabilities and prioritize retention efforts strategically. Additionally, real-time monitoring capabilities facilitate prompt responses to emerging churn risks, enhancing proactive retention strategies. Collaboration among stakeholders and continuous refinement of churn analysis methodologies ensure alignment with business objectives and evolving requirements. Ultimately, by harnessing the analytical power of Tableau and Power BI, banks can optimize customer retention strategies, reduce churn rates, and foster long-term customer relationships, thereby enhancing profitability and competitiveness in the banking industry.

REFERENCES:

1. Sarkaft Saleh1 · Subrata Saha1 Customer retention and churn prediction in the telecommunication industry:a case study on a Danish university Received: 27 February 2023 / Accepted: 16 May 2023 © The Author(s) 2023

2 Fathian M, Hoseinpoor Y, Minaei-Bidgoli B (2016) Offering a hybrid approach of data mining to predict thecustomer churn based on bagging and boosting methods. Kybernetes..

3. Wael Fujo S, Subramanian S, Ahmad Khder M (2022) Customer churn prediction in telecommunication industryusing deep learning. Inf Sci Lett 11(1):24.

4. Beeharry Y, Tsokizep Fokone R (2022) Hybrid approach using machine learning algorithms for customers' churnprediction in the telecommunications industry. Concur Comput: Pract Exp 34(4):e6627.

5. Idris A, Rizwan M, Khan A (2012) Churn prediction in telecom using Random Forest and PSO based databalancing in combination with various feature selection strategies. Comput Electric Eng 38(6):1808–1819.

6. Lalwani P, Mishra MK, Chadha JS, Sethi P (2022) Customer churn prediction system: a machine learningapproach. Computing 104(2):271–294

7 Bahri-Ammari N, Bilgihan A (2017) The effects of distributive, procedural, and interactional justice oncustomer retention: an empirical investigation in the mobile telecom industry in Tunisia. J Retail Consum Serv37:89–100..

8 Coussement K, Lessmann S, Verstraeten G (2017) A comparative analysis of data preparation algorithms forcustomer churn prediction: a case study in the telecommunication industry. Decis Support Syst 95:27–36

9 Alzubaidi AMN, Al-Shamery ES (2020) Projection pursuit Random Forest using discriminant feature analysismodel for churners prediction in telecom industry. Int J Electr Comput Eng.

10. Deng Z, Lu Y, Wei KK, Zhang J (2010) Understanding customer satisfaction and loyalty: an empirical study ofmobile instant messages in China. Int J Inf Manage 30(4):289–300

11. Farooq M, Raju V (2019) Impact of over-the-top (OTT) services on the telecom companies in the era oftransformative marketing. Glob J Flex Syst Manag 20(2):177–188.

12. Jyh-Fu Jeng D, Bailey T (2012) Assessing customer retention strategies in mobiletelecommunications: hybrid MCDM approach. Manag Decis 50(9):1570–1595

13. Praseeda CK, Shivakumar BL (2021) Fuzzy particle swarm optimization (FPSO) based featureselection and hybrid kernel distance based possibilistic fuzzy local information C-means (HKDPFLICM)clustering for churn prediction in telecom industry. SN Appl Sci 3:1

14 .Praseeda CK, Shivakumar BL (2021) Fuzzy particle swarm optimization (FPSO) based feature selectionand hybrid kernel distance based possibilistic fuzzy local information C-means (HKDPFLICM) clusteringfor churn prediction in telecom industry. SN Appl Sci 3:1

15. . Jensen D (2021) Telenor vil væk fra aggressiv fokus på priser - satser på nye forretningsområder.

16. Alalwan AA, Baabdullah AM, Rana NP, Tamilmani K, Dwivedi YK (2018) Examining adoption of mobileinternet in Saudi Arabia: extending TAM with perceived enjoyment, innovativeness and trust. TechnolSoc 55:100–110

17 Statista (2022) Telecommunications in Denmark

18. . Bahri-Ammari N, Bilgihan A (2017) The efects of distributive, procedural, and interactional justice on customer retention: an empirical investigation in the mobile telecom industry in Tunisia. J Retail Consum Serv 37:89–100.

19. Keramati A, Ardabili SM (2011) Churn analysis for an Iranian mobile operator. Telecommun Policy 35(4):344–356.

20. . Ullah I, Raza B, Malik AK, Imran M, Islam SU, Kim SW (2019) A churn prediction model using random forest:analysis ofmachine learning techniques for churn prediction and factor identifcation in telecom sector. IEEE Access7:60134–60149.

21. Bhattacharyya J, Dash MK (2022) What do we know about customer churn behaviour in the telecommunicationindustry? A bibliometric analysis of research trends, 1985–2019. FIIB Bus Rev 11(3):280–302.

22. Jaffry, S. W., & Ghulam, Y. (2019). Customer churn prediction in retail banking using data mining techniques.Journal of Financial Services Marketing, 24(4), 145-157.

23. Jolly K (2018) Machine learning with scikit-learn quick start guide: classification, regression, and clusteringtechniques in Python.

24 Wade C (2020) Hands-on gradient boosting with XGBoost and scikit-learn: perform accessible machine learningand extreme gradient boosting with Python. Packt Publishing Ltd. ISBN: 1839213809, 9781839213809

25. Baesens, B., Van den Poel, D., & Egmont-Petersen, M. (2020). An empirical assessment of kernel typeperformance for least squares support vector machine classifiers. European Journal of Operational Research, 159(2),
435-450.

26. Chae, B. K., & Ho, S. Y. (2018). A new application of the data mining technique for customer churn predictionand classification analysis in telecommunications industry. Computers & Industrial Engineering, 118, 260-272.

27. Miralaei, M. H., & Aalaei, A. (2019). A novel approach for customer churn prediction in mobiletelecommunications industry using deep belief networks. Expert Systems with Applications, 42(22), 882-891.

28. Bouguessa, M., & Bouatta, W. (2021). Mining electronic banking data for effective customer relationshipmanagement. Expert Systems with Applications, 38(10), 12533-12540.

29. Elahi, S., & Bahrami, M. A. (2019). Customer churn prediction in banking industry using a hybrid data miningapproach. Journal of Retailing and Consumer Services, 51, 36-44.

30. Lam, S. S. (2022). Predicting customer churn in banking industry and support decision making using logisticregression. Procedia Economics and