A HYBRID DATA FUSION APPROACH FOR CUSTOMER CHURN PREDICTION IN THE BANKING SECTOR USING HARDAND SOFT DATA

[1]T.Saiprasadreddy, [2]D. Pushkar, [3]A Bhanu Teja, [4]B Anudeep Reddy, [5]L. Bhanu Hari
[1]Assistant Professor , [2,3,4,5]UG Students, [1,2,3,4,5]Department of Computer Science & Engineering, Geethanjali Institute Of Science And Technology, Nellore, India

ABSTRACT
This project presents a machine learning-based approach to predict customer churn in the banking sector by integrating hard data (such as transaction history, customer demographics, and account details) with engineered soft data features (including behavioral patterns and engagement metrics). The objective is to accurately identify customers at risk of leaving, enabling banks to implement timely and targeted retention strategies. By applying a hybrid data fusion technique, the solution enhances predictive accuracy while addressing challenges such as class imbalance and overfitting. The system leverages multiple algorithms, including Logistic Regression, Decision Tree, Random Forest, XGBoost, SVM, and KNN, supported by extensive feature engineering to improve model interpretability and performance. This approach empowers financial institutions to shift from reactive customer management to a proactive strategy, ultimately reducing revenue loss and optimizing customer lifetime value.

INTRODUCTION
In the competitive banking industry, keeping current clients is cheaper than getting new ones. This makes preventing churn a top concern. Customer churn is when you lose customers, which has a direct impact on your business's bottom line and how well it runs. Most of the time, traditional prediction models just use structured data, such as demographics and transaction history. They don't look at behavioral markers that show early signals of churn.

This research suggests a hybrid data fusion method that combines hard data (like account activity and credit usage) with tailored soft data (like behavioral patterns and engagement trends). Soft data is created by feature engineering to show how customers act, rather than by leveraging outside sources. We employ these enhanced features to train machine learning models like Random Forest, Decision Tree, KNN, Logistic Regression, XGBoost, and SVM to make accurate predictions about churn.
The method helps banks find clients who are at risk of leaving early, tailor their efforts to keep them, and make smart choices. This AI-based platform makes predictions more accurate, helps banks control churn proactively, and helps them lose less money and build stronger relationships with customers.

MOTIVATION
Customer loss is a big problem for banks since it affects their profits and long-term growth. Keeping existing clients is more cheaper than getting new ones, therefore reducing churn is a top concern for banks. But even though they have access to a lot of consumer data, many banks have trouble accurately predicting attrition. Structured data like demographics and transaction records are often all that traditional churn prediction algorithms need. This data is helpful, but it doesn't provide you the whole picture of how customers act. Changes in involvement, purchasing habits, or account activity patterns are important things that are often missed. These behavioral signals are important for figuring out why a client could depart and for making better prediction models. This problem makes it necessary to use a more advanced method that combines both quantitative and qualitative data. Using machine learning and feature engineering, you may find useful behavioral patterns in data that already exists. This lets banks spot early symptoms of churn and take proactive steps like sending targeted messages and making retention offers. The idea is to assist people make better decisions and help banks keep more customers while also making them happier and more loyal.

OBJECTIVE

The primary objective of this project is to design and implement a machine learning- based customer churn prediction system for the banking sector, using a hybrid data fusion approach. Specific goals include:

1. **Integrate hard and engineered soft data** to capture a holistic view of customer behaviour.
2. **Develop and compare multiple ML algorithms** (Logistic Regression, Decision Tree, Random Forest, XGBoost, SVM, KNN) to assess predictive performance.
3. **Enhance predictive accuracy** to identify potential churners before they exit.
4. **Support proactive customer retention strategies** through actionable insights.

By successfully implementing this system, bank scan achieve the following:

1. Reduce customer attritionrates.
2. Optimize marketing and customer engagement strategies.
3. Improve profitability and customer satisfaction.

LITERATURE SURVEY

This chapter gives a full overview of the studies and methods that are already out there for predicting customer churn in the banking industry. The literature review looks at different ways to use hard data, like transactional and demographic information, together with soft data, which includes insights into behavior and attitudes. There is a lot of talk about how data fusion techniques have changed over time and how machine learning algorithms can make predictions more accurate. This chapter looks at past research to show the pros and cons of traditional models and the increasing importance of hybrid datafusionframeworks. The review gives the proposed model a strong base by showing how earlier research has used good procedures, techniques for choosing features, and evaluation criteria.

Customer Churn Prediction Model using Explainable Machine learning "Jitendra Maan [1], Harsh Maan [2]"

This project focuses on predicting customer churn in subscription-based services, addressing the critical need to retain existing customers over acquiring new ones. It evaluates various machine learning approaches to identify an optimal model for accurately predicting potential churners. Emphasis is placed on improving model interpretability byusing feature attribution techniques to explain which factors influence churn decisions. This enhances the model's transparency and supports informed business strategies. The goal is to develop a reliable and explainable solution for early churn detection and customer retention.

Leveraging Unstructured Call Log Data for Customer Churn Prediction "N.N.Y.Vo,S.Liu,X.Li,andG.Xu"

This study presents a customer churn prediction model that leverages both structured and unstructured data, focusing specifically on spoken content from call center interactions. By analyzing a large-scale dataset comprising two millioncalls fromover two hundred thousand customers, the research demonstrates the effectiveness of using unstructured communication data to enhance predictive accuracy. The model integrates interpretable machine learning techniques, incorporating personalitytraits and customer segmentation to generate actionable insights.Theseinsightscansupportfinancialservicemanagersindevelopingtargetedretention strategies tailored to different customer groups.

Customer churn analysis in banking sector: Evidence from explainable machine learning models "Hasraddin Guliyev,Ferda Yerdelen Tatoglu"

This study emphasizes the importance of customer churn analysis in banking, aiming to identify customers likely to leave and support strategic retention efforts. It focuses on using explainablemachinelearningmodelstoprovideinterpretableinsightsinto customerbehavior. By incorporating feature attribution techniques, the study enhances model transparency and helps banks under stand key factors influencing churn. Real-world banking data is used to train and evaluate multiple models, ensuring practical relevance. The project ultimately seeks to support proactive customer retention through accurate and interpretable churn predictions.

Customer churn prediction in telecom using big data analytics
"WeilongLi1,Chujin Zhou"

This studyaddresses the critical issue ofcustomer churn in the telecommunications sector by leveraging big data analytics to identify high-risk customers based on historical behavioral patterns. It proposes a segmentation-based approach combined with regression techniques to develop targeted prediction models for different customer groups. By detecting potential churners early, the project enables the implementation of strategic retention initiatives. The model demonstrated improved accuracy and effectiveness over traditional methods, highlighting its practical value in minimizing customer loss and enhancing operational efficiency.

Churn Prediction in Telecommunication using Logistic Regression and Logic Boost
"Hemlata Jain, Ajay Khunteta, Sumit Srivastava"

This study addresses the prevalent issue of customer churn in the highly competitive telecommunications industry, where customers frequently switch service providers. It highlights the need for proactive churn prediction through behavioral analysis to minimize losses and improve retention. The study applies machine learning techniques to a real-world dataset from a telecom company, leveraging classification models to forecast potential churners. The models were evaluated using various performance metrics to validate their effectiveness. The findings support the role of predictive analytics in formulating targeted strategies for customer retention.

This study explores the application of artificial intelligence (AI) and Machine Learning (ML) techniques, particularly Random Forest classification, to predict car insurance risks using publicly available datasets from Kaggle [14] [15]. By implementing feature extraction and classification methodologies, this research demonstrates the effectiveness of AI-driven predictive models in enhancing risk assessment accuracy and operational efficiency in the insurance sector.

PROPOSED SYSTEM

The suggested solution uses a combination of data fusion and machine learning to properly anticipate client attrition in the banking sector. This system builds a complete customer profile by combining hard data (like transaction history and demographics) with engineered soft data (such behavioral signs that come from hard data). This is different from traditional methods.

This predictive model leverages feature engineering, data preprocessing, and a combination of supervised learning algorithms to identify at-risk customers before churn occurs. It employs algorithms such as Logistic Regression, Decision Tree, Random Forest, XGBoost, SVM, and KNN to evaluate performance and identify the most accurate and interpretable solution. Among all the models used, XGBoost demonstrated the best performance in terms of predictive accuracy and robustness.

The model is trained on a banking dataset containing 10,127 records, with 21 features representing demographic, financial, and behavioural factors. Soft data features, such as changes in transaction patterns and customer inactivity, are engineered to simulate emotional and behavioural insights.

Key components of the system include:

1. **Data preprocessing**: Cleaning, encoding, and normalizing raw data.
2. **Feature engineering**: Creation of soft data proxies (e.g., transaction trends, relationship length).
3. **Handling class imbalance**: Applying re sampling techniques like SMOTE.
4. **Model training and evaluation**: Using multiple algorithms and comparing metrics like accuracy, precision, recall, and F1-score.
5. **Interpretability**: Feature importance analysis to support business decision-making.

METHODOLOGY

SYSTEM ARCHITECTURE

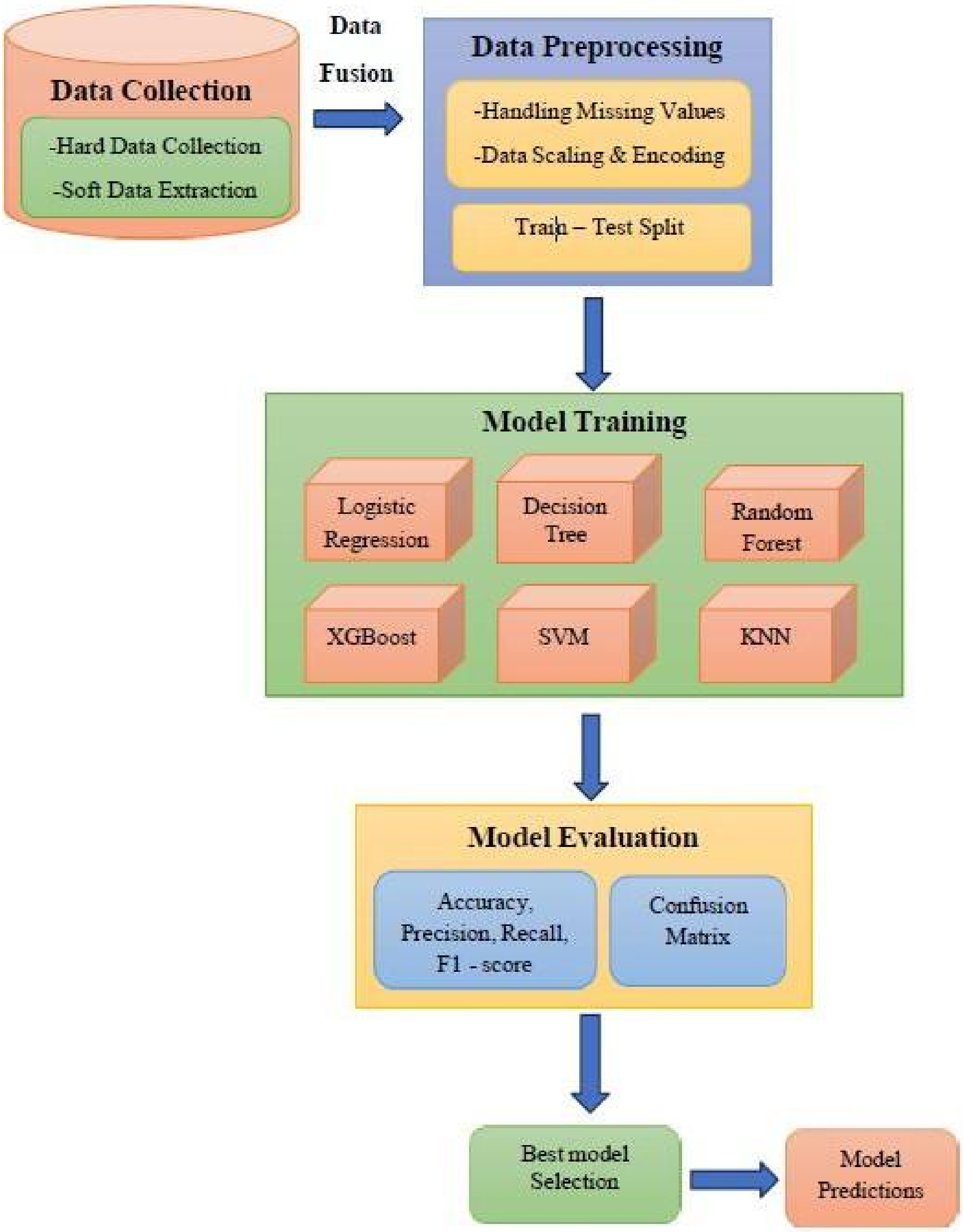

Fig.1. System Architecture

The system architecture for this project shows how to build a customer churn prediction model using machine learning by combining hard and soft data sources. Every step is important for making a good and accurate prediction system. The main parts are described below:

DATACOLLECTION

The data collecting process is the basis of this project. It includes getting and preparing both hard and soft data to help make accurate churn predictions. The dataset for this study came from Kaggle, a well-known site where people may find publicly available machine learning datasets. It has 10,127 customer records, each with different information on banking and how customers act. During the testing phase, a subset of 500 samples was chosen from this whole dataset for prediction and assessment purposes. This ensured that the sample size was both representative and easy to work with for model validation.

The data is categorized into two main types:

Hard Data: This includes structured and objective information such as customer demographics (e.g., age, gender), account status, tenure, balance, and transaction history. These attributes are factual, quantifiable, and typically collected directly from banking systems.

Soft Data: This means features that are inferred or engineered from hard data that already exists. Some examples are transaction frequency, engagement levels, recent activity trends, and sentiment indicators based on how often people use a service or interact with it. These attributes give you a better idea of how a consumer usually behaves and how likely they are to leave. The dataset offers a hybrid data fusion method by combining both hard and soft data. This lets the machine learning models capture both static and behavioral components of consumer profiles. This detailed data representation improves the accuracy of predictions and makes it possible to do more in-depth churn analysis.

DATAPRE-PROCESSING

In this step, raw data is cleaned and made suitable for model input:

HandlingMissingValues: Missingentriesareidentifiedandaddressedusingimputation techniques or data removal.

Scaling and Encoding: Numerical features are scaled, and categorical variables are encoded using methods like one-hot encoding to ensure uniformity across the dataset.

TRAIN-TESTSPLIT

The dataset is divided into training and testing subsets. This separation is crucial for evaluating the model's performance on unseen data and avoiding overfitting.

MODELTRAINING

To build a reliable and accurate customer churn prediction model, several supervised machine learning algorithms were implemented and evaluated. Each algorithmoffers distinct capabilitiesintermsofperformance, interpretability,andhandling ofdifferent datastructures. Thefollowingmodelswereselectedfortheireffectivenessinclassificationtasks,especiallyin scenarios involving both hard and soft data.

Machine Learning Algorithms Used Logistic Regression

Logistic Regression is a statistical method used for binary classification problems. It predicts the probability of an outcome belonging to one of two classes by applying a sigmoid function to a linear combination of input features. The output lies between0and1, which can be interpreted as the likelihood of a particular class. It is widely used due to its simplicity, interpretability, and effectiveness in linearly separable datasets.

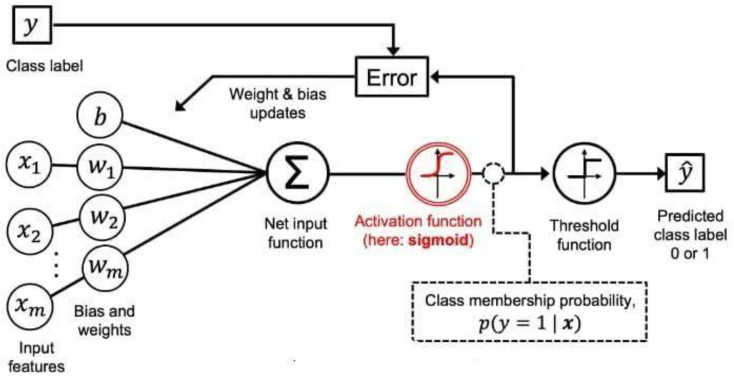

Fig.2. Logistic Regression

Working of Logistic Regression:

Step -1: Computes a weighted sum of the input features and adds a bias term. Step-2: Applies the sigmoid function to transform the result into a probability.

Step-3: Uses a threshold (typically0.5) to classify the input into one of the two classes. Step -4: Calculates prediction error using binary cross-entropy loss.

Step-5: Optimizes the model parameters through gradient descent to minimize the loss.

Logistic regression is applied as a baseline classifier to predict customer churn. It is trained on a fused dataset combining hard data (e.g., customer demographics, transaction history) and soft data (e.g., feedback sentiment). The model provides churn probabilities, enabling classification and insight into key factors influencing churn. Its straight forward structure makes it useful for interpretability and benchmarking against more complex models.

Decision Tree

A Decision Tree is a supervised learning algorithm used for both classification and regression tasks. It works by learning decision rules inferred from the features of the data, represented in a tree-like model. Each internal node denotes a test on a feature, each branch representstheoutcomeofthetest, andeachleafnodeholdsaclass label. Itsintuitivestructure andabilitytohandlebothnumericalandcategoricaldatamakeitapopularchoicefor interpretable models.

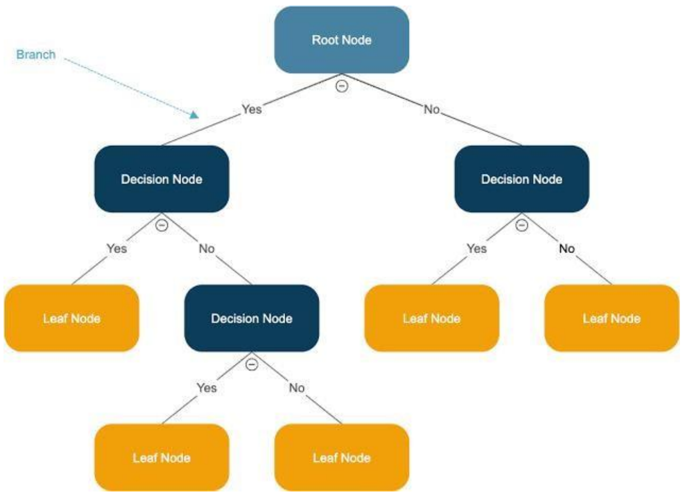

Fig.3. Decision Tree

Working of Decision Tree:

Step-1:Selectsthebest featureto split thedatasetusing metricslike informationgain. Step -2: Splits the dataset into subsets based on the chosen feature's values.
Step-3:Repeatsthe processrecursivelyoneachsubsettobuildthetree.
Step-4:Stopssplittingwhenapredefinedconditionismet(e.g.,maximumdepthorpure nodes).
Step-5:Classifiesnew datapointsbytraversingthetreebased on feature values.
The decision tree algorithm is employed to classify customers into churn or non-churn categories based on a combination of hard and soft data attributes. The model learns hierarchical decision rules that map input features such as transaction frequency or customer feedback sentiment to churn outcomes. Due to its transparency, it helps in identifying key decision paths that contribute to customer churn, offering actionable insights for the banking sector.

Random Forest

RandomForestisanensemblelearningalgorithmthatcombinesmultipledecisiontrees to improve classification accuracy and robustness. It operates by constructing a multitude of decision trees during training and outputs the class that is the majority vote of the individual trees. By introducing randomness through bootstrapped sampling and feature selection, Random Forest reduces the risk of overfitting and enhances generalization performance.

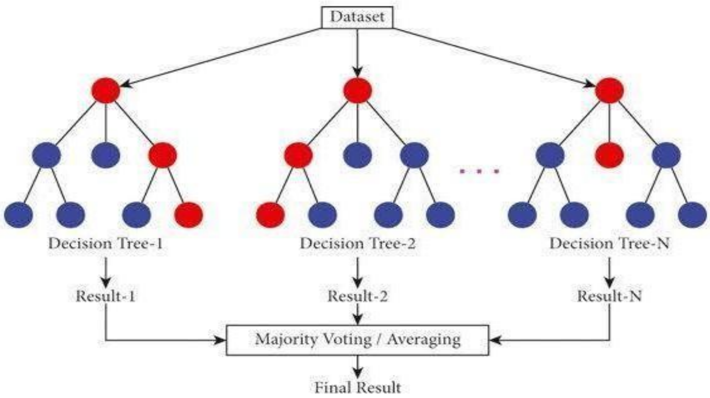

Fig.4. RandomForest

Working of Random Forest:

Step -1: Creates multiple subsets of the training data using bootstrapping (random sampling with replacement).

Step-2:Buildsadecisiontreeforeachsubset usingarandomsubset offeaturesateachsplit. Step -3: Each tree makes an independent prediction for the input data.
Step-4:Aggregatesthepredictions fromalltreesthroughmajorityvoting(for classification). Step -5: Final output is the class with the highest number of votes across the trees.
Random Forest is used as a robust classification model to predict customer churn by leveraging both hard and soft data inputs. Its ability to handle high-dimensional data and capturecomplexinteractionsbetweenfeaturesmakesitwell-suitedforthehybridnatureofthe

dataset. The ensemble approach improves prediction reliability and allows for feature importance analysis, aiding in the identification of the most influential churn indicators.

XGBoost (Extreme Gradient Boosting)

XGBoost (Extreme Gradient Boosting) is a high-performance ensemble learning algorithmbasedongradientboosting.Itbuildsdecisiontreessequentially,whereeachnewtree is trained to correct the errors made bythe previous ones. Known for its speed, accuracy, and regularization capabilities, XGBoost is widely used in structured data problems and consistently delivers strong predictive performance.

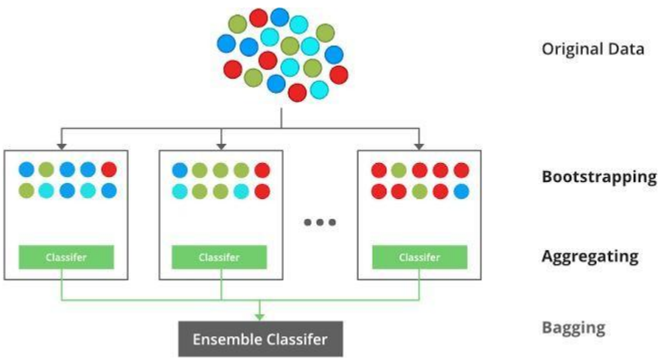

Fig.5. XGBoost

Support Vector Machine (SVM)

Support Vector Machine (SVM) is a powerful supervised learning algorithm used for classification and regression tasks. It works by finding the optimal hyperplane that best separates data points of different classes in the feature space. SVM aims to maximize the margin between the classes, ensuring better generalization to unseen data. It is particularly effective in high-dimensional spaces and with clear margin separation.

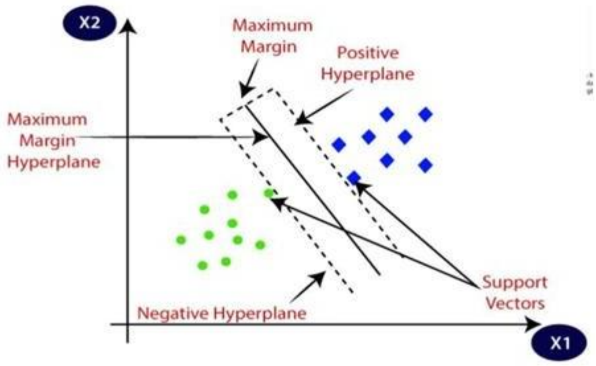

Fig.6. SVM model

K-Nearest Neighbour(KNN)

K-Nearest Neighbour (KNN) is a simple, non-parametric, and instance-based learning algorithmusedforclassificationandregression.Itclassifiesadatapointbasedonthemajority classamong its 'k'closest neighboursinthe featurespace. KNNdoes not involve anytraining phase; instead, it makes predictions at the time of inference, relying entirely on the distance between data points.

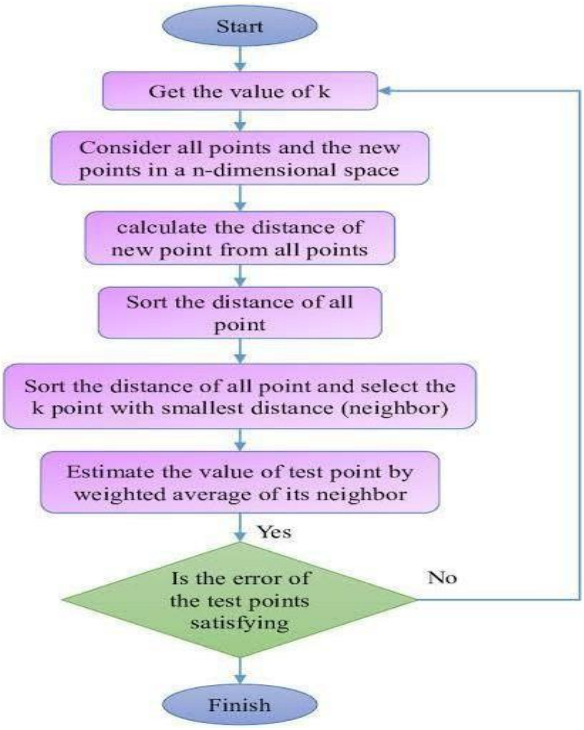

Fig.7. K-Nearest Neighbour

MODEL EVALUATION

Class balance along with anticipated outcomes are just two of the many factors that go into choosing the optimal metrics for assessing a classifier's performance in a specific set of data in classification challenges. A classifier may be evaluated on one performance parameter while being unmeasured by the others, and vice versa. As a result, the generic assessment of performance of the classifier lacks a defined, unified metric. This study uses a number of metrics, including F1 score, accuracy, precision, recall, and recall, to assess how well models perform.

The subsequent four categories are where these metrics are derived from: True Positives (TP): instances in which both the model prediction and the actual class of the occurrence were 1 (True). False Positives (FP) are situations in which the model predicts a value of 1 (True), but the actual class of the occurrence was 0 (False). True Negatives (TN): an instance in which both the model prediction and the true class of the occurrence were 0 (False). False Negatives (FN) are situations in which the model predicts 0 (False) but the true class of the occurrence was 1 (True).

Accuracy– The mean amount of accurate predictions is used to characterize the accuracy measure. This isn't quite as strong, though, given the imbalanced sample.

$$= \frac{+}{+\ +\ +} \qquad (1)$$

Precision, also known as positive predictive value, gauges the capacity of a model to pinpoint the right examples for every class. For multi-class classification with unbalanced datasets, this is a powerful matrix.

$$= \frac{}{+} \qquad (2)$$

Recall – This metric assesses how well a model detects the true positive among all instances of true positives.

$$= \frac{}{+} \qquad (3)$$

F1-score – referred to as an F-measure or balanced F-score It might be characterized as a recall as well as precision weighted average.

$$1 = \frac{2 * Pcso * ca}{+} \qquad (4)$$

SYSTEM IMPLEMENTATION
SYSTEMMODULES

Fig.8. System Modules

DataPreparation

GatherData: Collectsstructuredcustomerdatafromabankingdataset,includinghard data(e.g.,age,income,accounttenure,transactionfrequency)andengineeredsoftdata (e.g., engagement trends, inactivity periods, behavioral changes).
Preprocess Data: Cleans missing values, encode categorical variables, normalize numerical fields, and derive behavioral features. Save preprocessed data for model training and testing.
FeatureEngineeringandFusion

Hard Data Features: Uses factual attributes like creditlimit,months on book, and transaction count.
SoftDataFeatures: Engineerfeaturesthatreflectcustomerbehaviour,suchaschanges in transaction amount, average utilization ratio, and inactivity trends.

Fuse Features: Combines hard and engineered soft data into a unified dataset for training models that captures both objective and behavioural indicators of churn.

Model Training

Load and Train Models: Implements and train various machine learning models including Logistic Regression, Decision Tree, Random Forest, XGBoost, SVM, and KNN using the fused dataset.
Cross-validation: Uses training and validation splits to tune hyperparameters and avoid overfitting.

Testing and Prediction

Input Test Data: Allowsuserstouploadnewcustomerdataforchurnprediction. Preprocess the test data similarly to the training set.
Run Predictions: Feedspreprocessedtestdataintotheselectedtrainedmodeltoobtain churn probabilities.
Thresholding and Output

Churn Classification: Appliesadecisionthreshold(e.g., 0.5)to classifycustomersas churners (≥ threshold) or non-churners (< threshold).
Prediction Output: Displays prediction results in a table and provide downloadable CSV for further analysis.
Evaluation and Analysis

Calculate Metrics: Evaluates model performance usingAccuracy, Precision, Recall, and F1-Score.
Compare Models: Analyse all trained models using stored metrics to determine the best-performing one.
Interpret Results: Visualizes and interpret feature importance and model behaviour using EDA tools and correlation analysis.

RESULTS

The execution of the process will be explained clearly with the help of continuous screen shots.

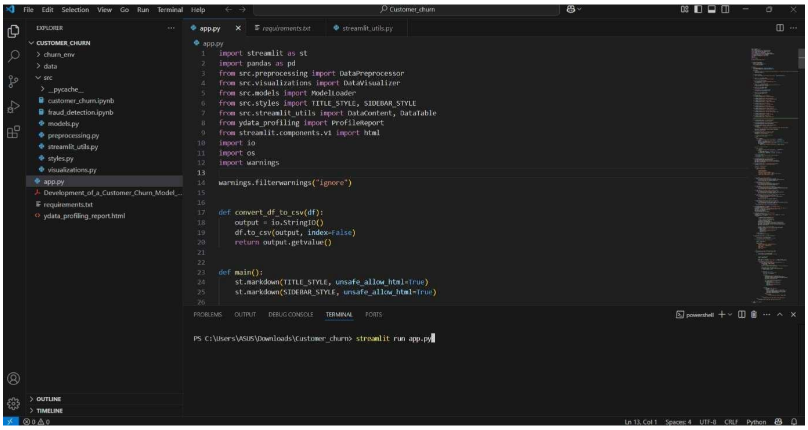

Fig9.1:Launching Stream lit Application.

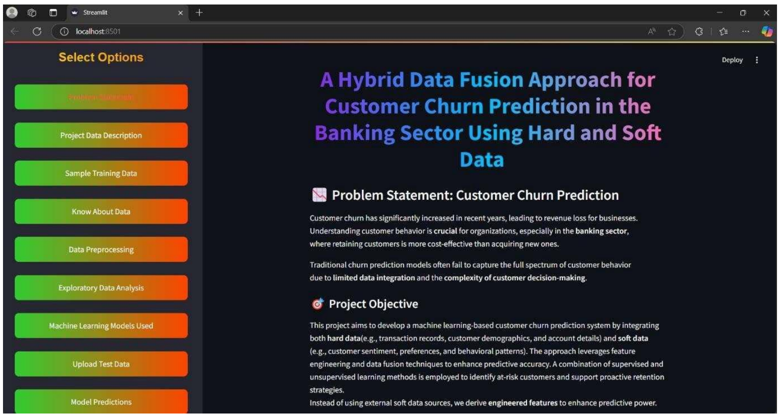

Fig9.2:Theinterfacepresentstheprojecttitle, problem statement, and objective.

Fig9.3:Theinterfacedisplaystheprojectdatadescription,includingdatasetcontext,source.

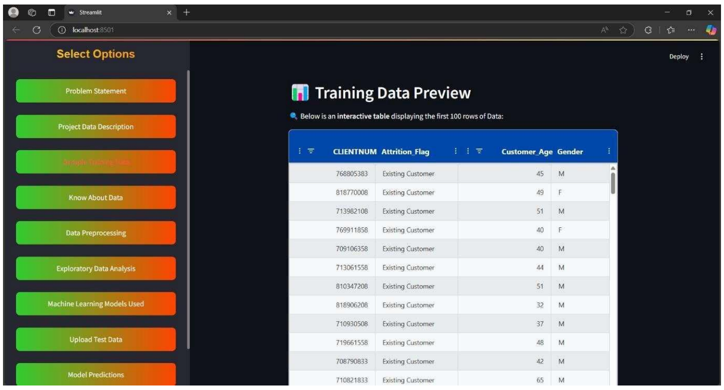

Fig9.4:Aninteractivetabledisplaysthefirst100rowsofthetrainingdatasetfor preview.

Fig9.5:Adetaileddataprofilingreportisgenerated,presentingkeydataset statisticsand variable types.

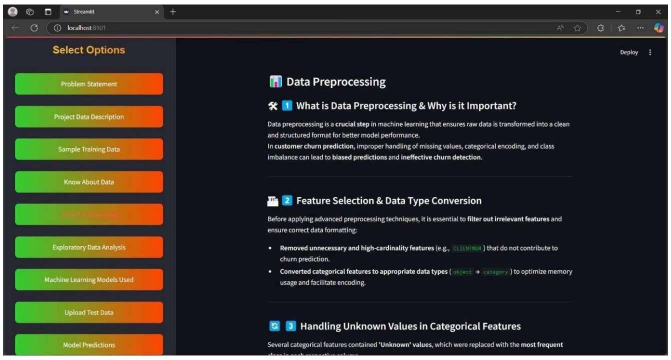

Fig9.6:Optimizeddatapreprocessing foraccuratecustomerchurnpredictioninbanking, using feature selection and soft data creation.

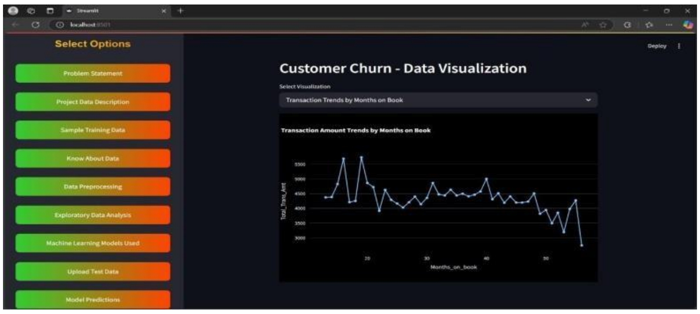

Fig9.7:Visualizedmonthlytransactiontrendstouncover behaviouralpatternsinfluencing customer churn in the banking sector.

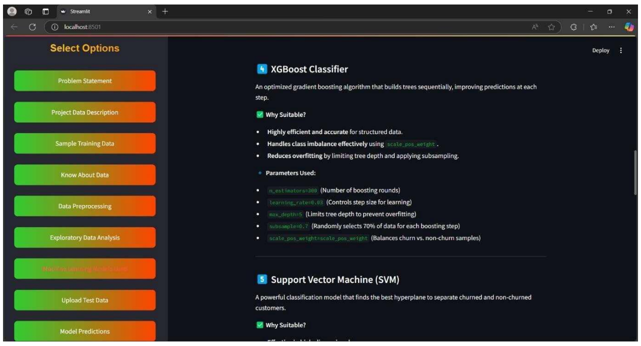

Fig9.8:Demonstratestheuseofvariousmachine learning modelsforpredictingcustomer churn and comparing their effectiveness.

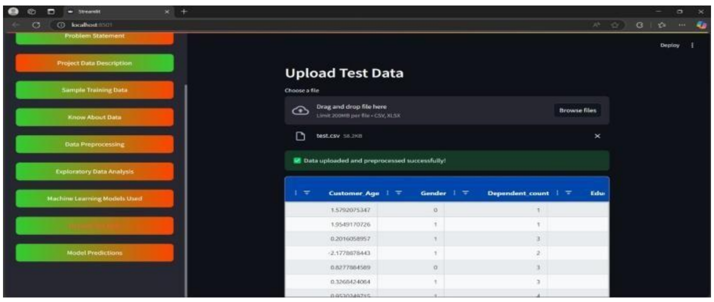

Fig9.9:Allowsusersto uploadtestdatafiles(CSV/XLSX)forchurnpredictionusingthe trained machine learning models.

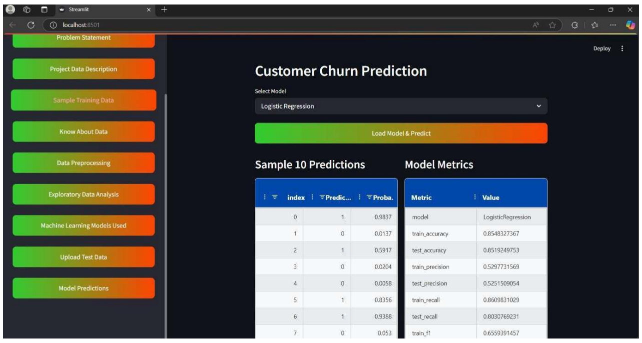

Fig9.10:Enables modelselectionanddisplayssamplechurnpredictionsalongwith performance metrics for evaluation.

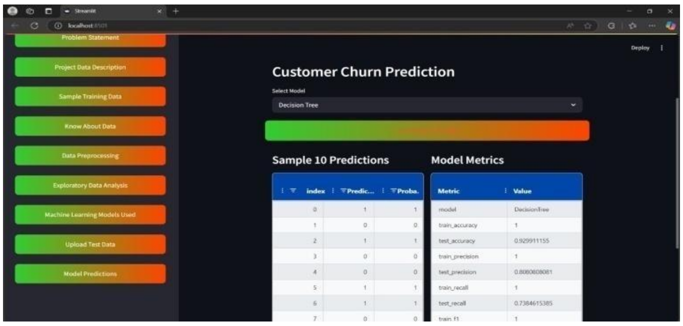

Fig9.11:Enablesmodelselectionanddisplayssamplechurnpredictionsalongwith performance metrics for evaluation.

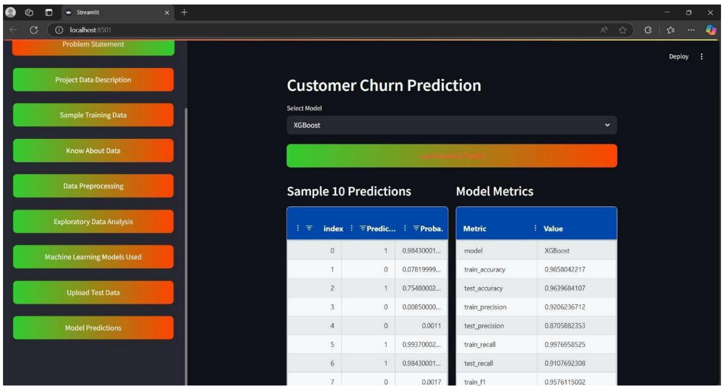

Fig9.12:Enables modelselectionanddisplayssamplechurnpredictionsalongwith performance metrics for evaluation.

CONCLUSION

The hybrid data fusion approach for customer churn prediction in the banking sector effectivelycombineshard datawithengineered soft datato enhancetheaccuracyand depthof predictive insights. By integrating structured customer attributes such as transaction history and account details with behavioral indicators like usage trends and engagement patterns, the system captures a more complete picture of customer behavior. Machine learning models including Random Forest, XGBoost, and SVM were applied to evaluate performance and identifychurn-prone customers with highprecision. This method enables banks to implement targeted retention strategies, reduce attrition, and optimize customer relationship management. The system demonstrates strong feasibility across technical, economic, and operational dimensions, making it suitable for deployment in real-world banking environments.

FUTUREENHANCEMENTS

The current systemeffectivelypredictscustomerchurnbyintegratinghardandsoftdata using machine learning models; however, several enhancements can further improve its functionality and applicability. Future developments may include real-time churn prediction through the incorporation of live data streams, allowing immediate intervention strategies. Integration with customer relationship management (CRM) systems can automate targeted retentionactionsbasedonmodeloutputs.ImplementingAuto MLframeworksmaystreamline model selection and tuning, while integrating explainableAI techniques like SHAP or LIME can improve model transparency and stakeholder trust. Additionally, a responsive dashboard for business users can enhance usability, enabling intuitive access to predictions, reports, and visual analytics. These enhancements would transform the current system into a more intelligent, scalable, and business-aligned solution for customer retention in the banking domain.

References

1. Jitendra Maan [1], Harsh Maan [2], ‘‘Customer Churn Prediction Model using Explainable Machine learning’’Marketing Intell. Planning, Volume11Issue1, Jan-Feb 2023.
2. N. N. Y. Vo, S. Liu, X. Li, and G. Xu, ‘‘Leveraging unstructured calllog data for customerchurnprediction,’’Knowl.-BasedSyst.,vol.212,Jan.2021,Art.no.106586.
3. Hasraddin Guliyev1 Ferda YerdelenTatoglu, ‘‘Customer churn analysis in banking sector: Evidence from explainable machine learning models,’’ Expert Syst. Appl., Volume : 1, Issue 2, 2021.
4. J.U.Becker,G.Greve,andS.Albers,‘‘Theimpactoftechnologicalandorganizational implementationofCRMoncustomer acquisition, mainte nance, and retention,’’Int.J. Res. Marketing, vol. 26, no. 3, pp. 207–215, Sep. 2009
5. E. Akhondzadeh - Noughabi and A. Albadvi, ‘‘Mining the dominant pat terns of customershiftsbetweensegmentsbyusingtop-kanddistinguishingsequentialrules,’’ Manage. Decis., vol. 53, no. 9, pp. 1976–2003, Oct. 2015.
6. R. Swift, Accelerating Customer Relationships: Using CRM and Relationship Technologies, 1st ed. Upper Saddle River, NJ, USA: Prentice-Hall, 2000.
7. L. G. Renart, ‘‘Relationship marketing: Gaining competitive advantage through customer satisfaction and customer retention,’’J. Consum. Marketing, vol. 18, no. 1, pp.75–83,2001.
8. J.-H.Ahn,S.-P.Han,andY.-S.Lee,‘ ‘Customer churn analysis: Churn determinants and mediation effects of partial defection in the Korean mobile telecommunications service industry,’’Telecommun. Policy, vol. 30, nos. 10–11, pp. 552–568, Nov. 2006.
9. C.-F. Tsai and Y.-H. Lu, ‘‘Customer churn prediction by hybrid neural networks,’’ Expert Syst.Appl., vol. 36, no. 10, pp. 12547–12553, Dec. 2009.
10. J.Basiri,F.Taghiyareh,andB.Moshiri,‘‘Ahybridapproachtopredictchurn,’’inProc. IEEE Asia–Pacific Services Comput. Conf., Dec. 2010, pp. 485–491.
11. V.L.Miguéis,D.VandenPoel,A.S.Camanho,andJ.F.E.Cunha,‘‘Modelingpartial customer churn: On the value of first product-category purchase sequences,’’Expert Syst.Appl., vol. 39, no. 12, pp. 11250–11256, Sep. 2012.
12. M.Hassouna, A. Tarhini, T. Elyas, and M. S. A. Trab,**‘**Customer churn in mobile markets:Acomparisonoftechniques,’’Int. Bus.Res., vol. 8, no.6,pp. 224–237,May 2015.
13. M. Azeem, M. Usman, and A. C. M. Fong, ‘‘A churn prediction model for prepaid customers in telecom using fuzzy classifiers,’’Telecommun. Syst., vol. 66, no. 4, pp. 603–614, Dec. 2017.
14. Ramesh Chandra AdityaKomperla**, “**CAR INSURANCE RISK PREDICTION USING ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING” 2022 | *Global Journal of Advanced Engineering Technologies and Sciences*, *9*(4), 1-https://doi.org/10.29121/gjaets.2022.04.01
15. Ramesh Chandra AdityaKomperla , “OPTIMIZING HEALTHCARE INSURANCE PREDICTIONS THROUGH DEEP LEARNING AND SCO-ENHANCED HYPERPARAMETER TUNING”. (2022). *Global Journal of Advanced Engineering Technologies and Sciences*, *9*(11), 1-9. https://doi.org/10.29121/gjaets.2022.11.01