International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Customer Churn Prediction System Using Machine Learning

Mr. M. Hari Krishna[1], M. Sai Krishna Yadav[2], A Shivamani[3], Yokshitha[4], CH Vikas[5]

[1] Assistant Professor, Dept. of CSE-Data Science, ACE Engineering College, India
[2345] B.Tech CSE-Data Science, ACE Engineering College, India
Emails: harivenimallu@gmail.com, msaikrishnayadav3@gmail.com, shivamaniathineni10@gmail.com, polayokshitha@gmail.com, vv871590@gmail.com

A B S T R A C T

All over the world, in different sectors churn prediction plays a very important role in the growth of the organization. For the company's revenue and profit, customer churn is very harmful. The most important step to avoid churn is to detect churn and its reason.

The objective of this project is to predict the churn in organization sectors, by using well known Machine Learning techniques like SVM, Decision tree. The classification model will be built by analyzing historical data and then applying the prediction model based on the analysis. Predicting the customer churn rate will help the organization in knowing which category of customers generally tend to leave the organization. In this project finding the major factors that affect the customers to churn and analyze them by using Machine learning algorithms. Then churn gives the information of how many existing customers tend to leave the business, so lowering churn has an immense positive impact on the revenue streams. The project aims to provide actionable insights to help businesses reduce churn rates and improve customer satisfaction through data-driven strategies.

Keywords: Customer Churn, SVM (Support Vector Machine), Decision Tree, Churn Rate.

1. Introduction

The banking sector has long been one of the most prominent industries. Companies continuously strive to survive in this highly competitive market by implementing strategic approaches. Over time, it has been observed that retaining an existing customer is significantly more cost-effective than acquiring a new one. Organizations across the globe recognize customer churn that is, customers moving from one provider to another as a major loss, especially considering investments already made to attract those customers. This is one of the primary reasons why customer retention is so valuable for businesses. To address this challenge, many companies invest in customer churn prediction. This involves identifying which customers are likely to leave, allowing companies to take proactive measures to retain them. In this project, I focused on evaluating and analyzing a company's dataset using the decision tree, SVM machine learning algorithm. The goal was to predict whether a customer will churn and to assess the overall customer churn rate in the banking sector.

2. Literature Review

Lalwani (2025) collated and analyzed 30 studies (2021–2025) focused on machine learning models (Decision Tree, Random Forest, SVM, Gradient Boosting, Neural Networks). Ensembles and hybrid models consistently achieved the highest performance. The author emphasizes aligning model selection with specific industries.

Bhatnagar & Srivastava (2025) provided a telecom-focused review, highlighting challenges of data sparsity and class imbalance. They noted that ensemble and deep learning methods outperform traditional statistical models, reinforcing the need for continuous feature selection and model adaptation.

The literature demonstrates a clear shift toward ensemble and deep learning approaches, coupled with advanced feature engineering and imbalanced data handling. Emerging trends highlight the integration of interpretability (e.g. SHAP), profit-centric metrics, and temporal modelling to align technical performance with business goals. To contribute novel insights, future work could explore cross-domain applications, real-time churn detection frameworks, and holistic pipelines balancing accuracy, explainability, and profit.

3. Methodology

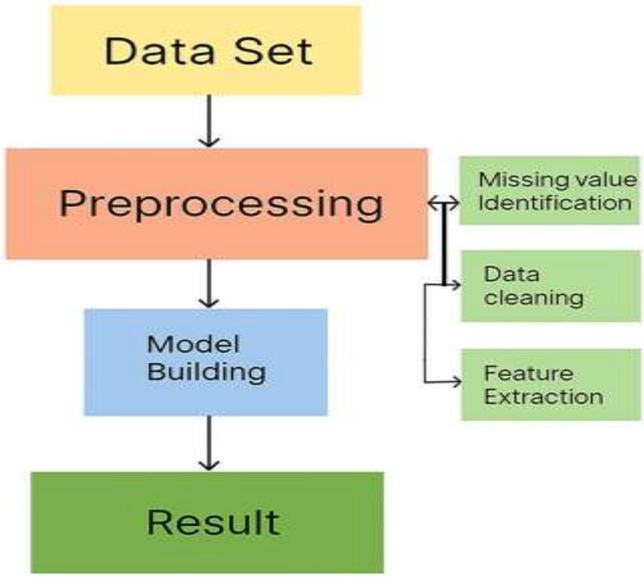

Fig. 1.Methodology.

The proposed approach for customer churn prediction follows a structured pipeline as shown in Figure 1:

1. **Dataset Acquisition**: Historical customer data is collected, containing relevant attributes such as demographics, usage behavior, and churn labels.
2. **Preprocessing**: This step includes:
   - **Missing Value Identification**: Detecting and addressing incomplete data.
   - **Data Cleaning**: Removing duplicates, correcting inconsistencies, and imputing missing values.
   - **Feature Extraction**: Selecting or engineering informative features to enhance model performance.
3. **Model Building**: Machine learning algorithms (e.g., SVM, Decision Tree) are trained on the processed data to classify churn vs. non-churn customers.
4. **Result Generation**: The final model is used to predict churn on new data, providing insights into churn drivers and aiding in customer retention strategies.

3.1 System Architecture

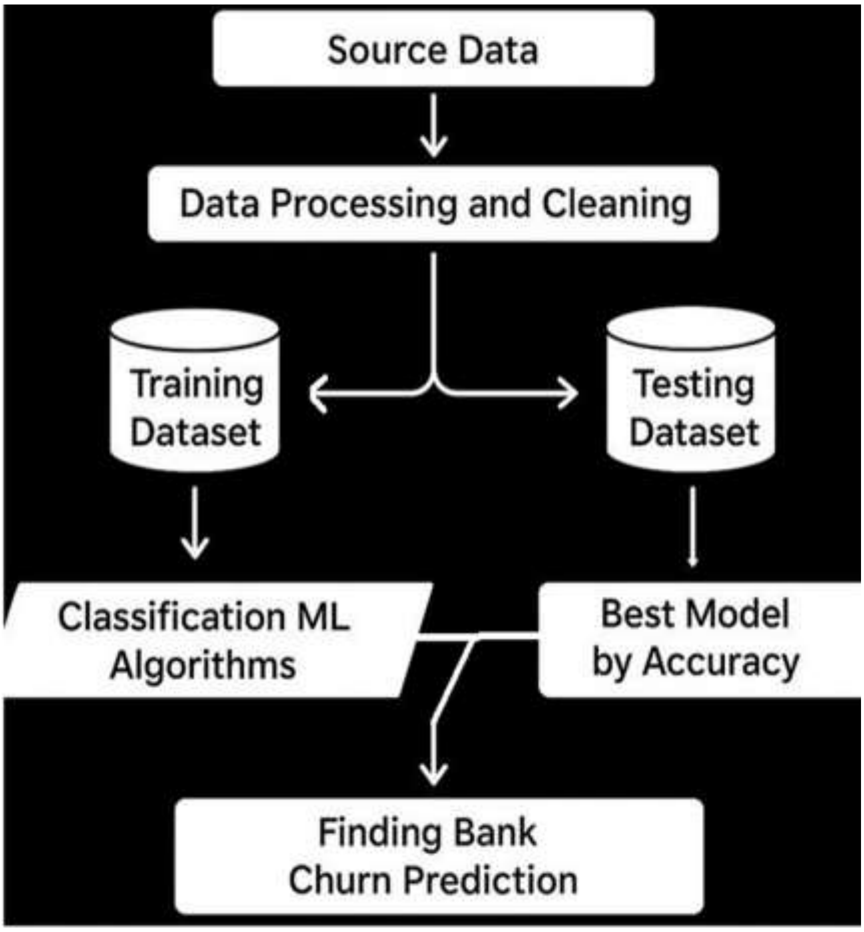

Fig. 2. Customer churn System Architecture

The proposed churn prediction system is structured into six key components:

1. **User Interface (UI)**: A front-end platform for users (e.g., bank staff) to input customer data (age, tenure, credit score, etc.) and view the churn prediction results.
2. **Input Validation Module**: Validates form inputs to ensure data completeness and correct formatting before processing.
3. **API Layer**: A Flask-based backend interface that transmits validated inputs to the prediction engine and returns the output to the UI.
4. **Prediction Engine**: A pre-trained machine learning model that receives processed inputs and returns churn predictions with associated probabilities.
5. **Model Wrapper**: Encapsulates model loading and ensures input-output consistency, bridging the API and the model.
6. **Interface Module**: Handles any required runtime preprocessing (e.g., encoding, scaling) to align real-time inputs with the model's expected format.

4. Output Screens:

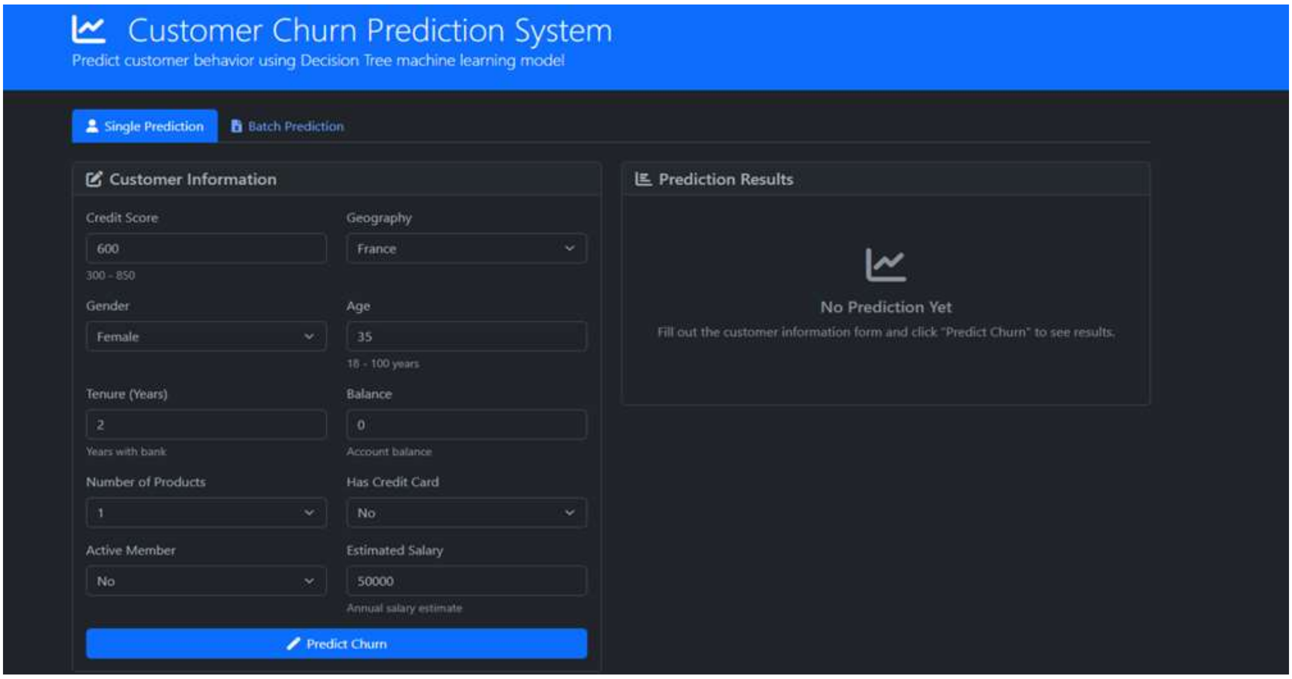

Fig 3.User Interface

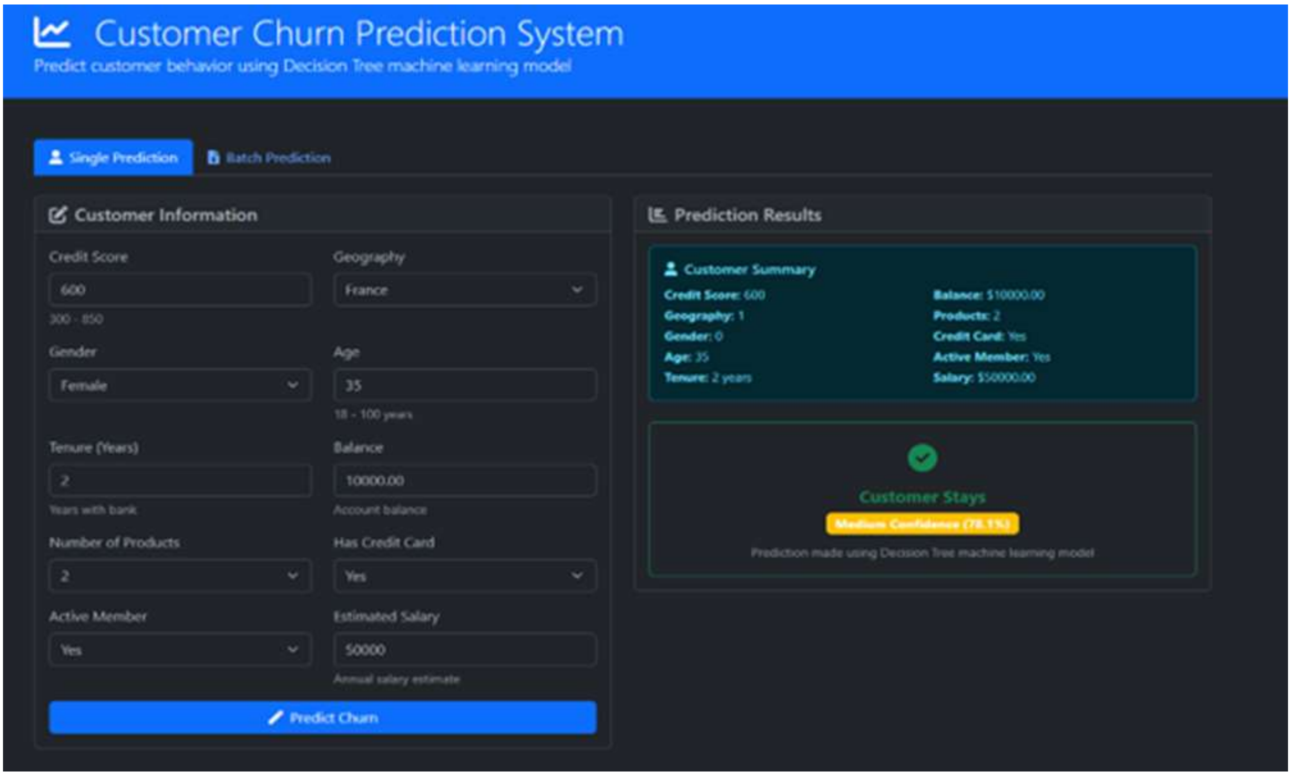

Fig 4.Sample output

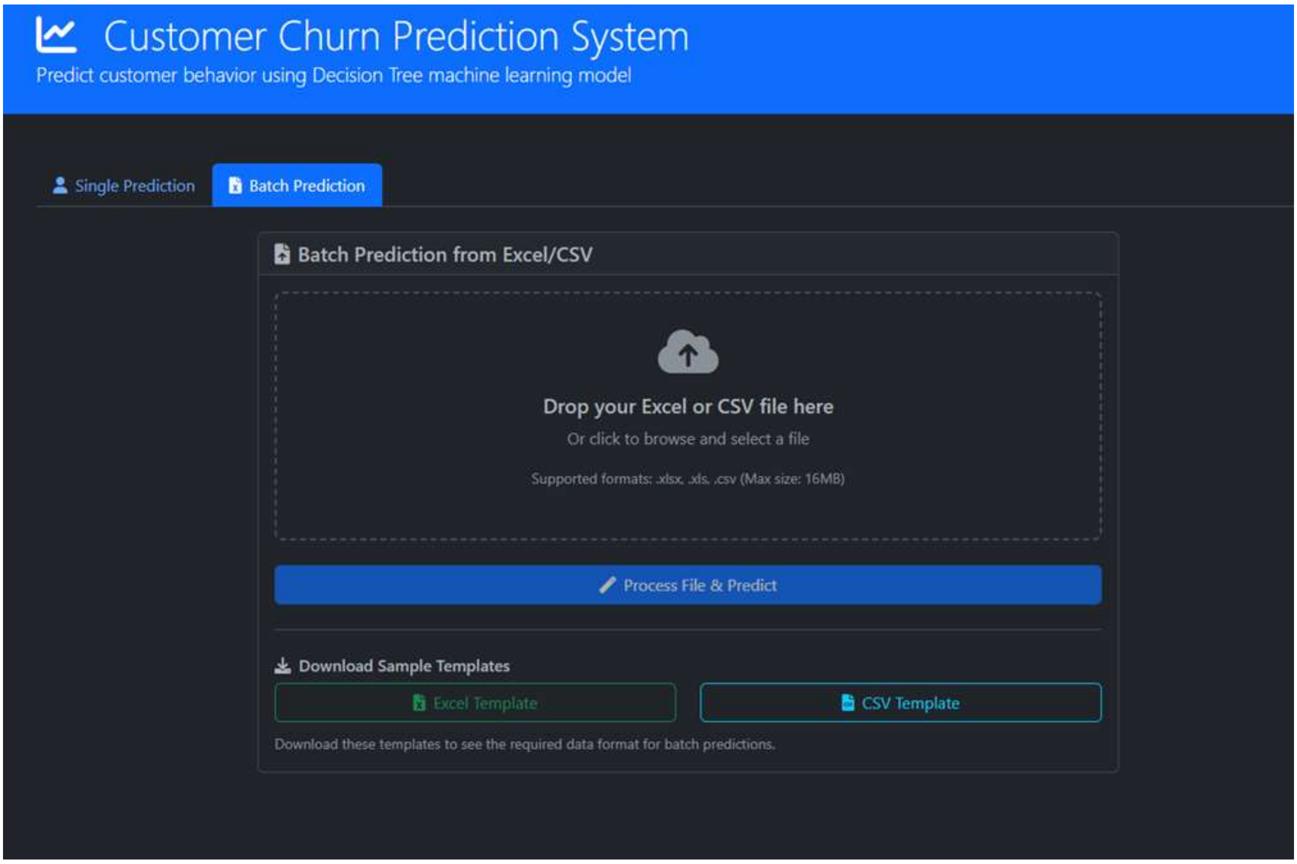

Fig 5. Batch Prediction

CreditScor	Geography	Gender	Age	Tenure	Balance	NumOfProducts	HasCrCard	IsActiveMember	EstimatedSalary	Decision_Tree_F	Prediction_Co	Prediction_Probability
619	France	Female	42	2	0	1	1	1	101348.88	Customer Exits	Medium	68.8
608	Spain	Female	41	1	83807.86	1	0	1	112542.58	Customer Exits	Low	53.8
502	France	Female	42	8	159660.8	3	1	0	113931.57	Customer Stays	High	85.7
699	Germany	Female	39	1	0	2	0	0	93826.63	Customer Exits	Medium	66.7
850	Spain	Male	43	4	125510.8	1	1	1	79084.1	Customer Stays	High	100
645	Germany	Male	44	8	113755.8	2	1	0	149756.71	Customer Stays	High	85.7
822	France	Male	50	7	0	2	1	1	10062.8	Customer Stays	High	80
376	Germany	Female	29	4	12551.72	4	1	0	119346.88	Customer Exits	High	90.9
501	Spain	Male	44	4	142051.1	2	0	1	74940.5	Customer Stays	High	100
684	France	Male	27	2	134603.9	1	1	1	71725.73	Customer Stays	Medium	70

Fig 6.Excel Results

5. Work Flow:

1. User Input via UI

- The user (e.g., a bank employee) opens the web/mobile application.
- They enter customer details such as age, balance, tenure, number of products, credit score, etc., into a form.

2. Input Validation

- The Input Validation Module checks if all fields are filled correctly (e.g., numeric fields contain numbers, no missing values). Or can be provided in a .csv, .xlxs file.
- If the input is invalid, the system prompts the user to correct the errors before proceeding.

3. Send Data to Backend (API Layer)

- Once input is valid, the UI sends the customer data to the backend server via a POST request (using REST API or similar).

4. Preprocessing (if needed)

- The backend checks if any runtime preprocessing (e.g., one-hot encoding, feature scaling) is needed.

- Preprocessed data is then passed to the model wrapper.

5. Model Prediction

- The pre-trained machine learning model receives the input.
- The model processes the data and predicts whether the customer is likely to *stay* or *leave* (churn).

6. Interpreting Results

- The output from the model is converted into a user-friendly format.
- Example: "Customer is likely to *churn* (78% probability). Recommended action: *Offer loyalty benefits*."

7. Return Prediction to UI

- The interpreted prediction is sent back to the UI via the API.

8. Display Output

- The UI displays the prediction and any additional suggestions to the user.

6. Results:

To enhance predictive performance, this study integrates Decision Tree and Support Vector Machine (SVM) classifiers using an ensemble learning approach. Decision Trees offer interpretability and effectively capture nonlinear relationships, while SVMs perform well in high-dimensional spaces with clear margin-based separation. By combining these models through ensemble techniques such as voting or stacking, the system benefits from the complementary strengths of both algorithms. As a result, the ensemble model achieved an improved accuracy of 0.864, outperforming the individual models and demonstrating greater reliability and generalization in customer churn prediction.

To evaluate the effectiveness of the churn prediction model, standard binary classification metrics were employed: **Accuracy**, **Precision**, and **Recall**.

- **Accuracy** measures the overall correctness of the model, defined as the ratio of correctly predicted instances (true positives and true negatives) to the total instances.
- **Precision** quantifies the proportion of correctly predicted positive cases out of all predicted positives, reflecting the model's ability to avoid false alarms.
- **Recall** (Sensitivity) assesses the model's ability to identify actual churn cases, calculated as the ratio of true positives to all actual positives.

For a test set of **10,000 instances**, the model achieved the following performance:

- **True Positives (TP)**: 5226
- **True Negatives (TN)**: 3414
- **False Positives (FP)**: 438
- **False Negatives (FN)**: 922

Metric Results:

- **Accuracy** = (5226 + 3414) / 10,000 = **0.864**
- **Precision** = 5226 / (5226 + 438) ≈ **0.923**
- **Recall** = 5226 / (5226 + 922) ≈ **0.850**

These results indicate that the model delivers high precision and strong recall, suggesting it is both accurate in identifying churn and reliable in minimizing false predictions.

7. **Conclusion:**

This study presents a robust and efficient machine learning framework for predicting customer churn, integrating Decision Tree and Support Vector Machine (SVM) models through ensemble learning techniques. The system is designed with a user-friendly interface, real-time prediction capabilities, and automated input validation, making it practical for deployment in business environments such as banking or telecom sectors. By leveraging the strengths of both Decision Tree and SVM classifiers, the ensemble model demonstrated improved predictive performance, achieving an accuracy of **0.864**, along with high precision and recall. These results highlight the effectiveness of ensemble-based approaches in enhancing model reliability and reducing customer attrition. Future work can explore integrating additional classifiers, implementing explainable AI components, and applying the system to real-time streaming data for continuous churn monitoring and proactive retention strategies.

8.Acknowledgement:

We express our sincere gratitude to all who supported us throughout the development of this project. We are especially thankful to Prof. Y. V. Gopala Krishna Murthy, General Secretary, and Mrs. M. Padmavathi, Joint Secretary, for providing us the opportunity and environment to carry out this work. Our heartfelt thanks to Dr. P. Chiranjeevi, Head of the Department, for his guidance and encouragement. We are deeply grateful to our internal guide Mr. M. Hari Krishna, and project coordinator Mr. G. Parwateeswar, for their consistent support, valuable feedback, and motivation throughout the project.

Lastly, we thank all the faculty members, staff for their constant encouragement and support.

9.References:

[1] **Verbeke, W., Martens, D., Mues, C., & Baesens, B.** (2012).

Building comprehensible customer churn prediction models with advanced rule induction techniques.

Expert Systems with Applications, 38(3), 2354–2364.

https://doi.org/10.1016/j.eswa.2010.08.098

[2] **Keramati, A., Jafari-Marandi, R., Aliannejadi, M., Ahmadi, A., & Mozaffari, M.** (2014).

Improved churn prediction in telecommunication industry using data mining techniques.

Applied Soft Computing, 24, 994–1012.

https://doi.org/10.1016/j.asoc.2014.08.041

[3] **Burez, J., & Van den Poel, D.** (2009).

Handling class imbalance in customer churn prediction.

Expert Systems with Applications, 36(3), 4626–4636.

https://doi.org/10.1016/j.eswa.2008.05.027