Performance Comparison of Random Forest and Support Vector Machine for Customer Churn Prediction

Bincy Mol B[1], Bhuvaneswari R[2]
[1,2]*Assistant Professor, Department of Computer Application, Hindustan College of Arts & Science, Padur, Chennai,India.*

***Abstract*—Customer churn prediction plays a crucial role in helping organizations retain valuable customers and reduce revenue loss. Accurate identification of customers who are likely to discontinue services enables businesses to implement effective retention strategies. This research presents a comparative analysis of two widely used machine learning algorithms, Random Forest (RF) and Support Vector Machine (SVM), for customer churn prediction using the Telco Customer Churn Dataset. The dataset undergoes comprehensive preprocessing, including missing value treatment, categorical data encoding, feature scaling, and feature selection to improve model performance. Both algorithms are trained and tested on the processed dataset, and their predictive capabilities are evaluated using Accuracy, Precision, Recall, and F1-Score metrics. Experimental results indicate that Random Forest achieves higher prediction accuracy and better overall classification performance due to its ensemble learning approach and ability to handle complex feature interactions. Support Vector Machine also demonstrates competitive performance by effectively identifying decision boundaries between churn and non-churn customers. The findings highlight the strengths and limitations of both algorithms and suggest that Random Forest is more suitable for customer churn prediction tasks in telecommunication environments. This research contributes to the development of intelligent customer retention systems and provides guidance for selecting appropriate machine learning models for churn management applications.**

***Index Terms*—Customer Churn Prediction, Random Forest, Support Vector Machine (SVM), Machine Learning, Classification, Predictive Analytics, Customer Retention, Telco Customer Churn Dataset, Data Mining**

I. INTRODUCTION

Customer retention has become a fundamental objective for modern business organizations operating in highly competitive markets. Retaining existing customers is generally more cost-effective than acquiring new ones, as loyal customers contribute significantly to long-term revenue generation and business growth. Organizations invest substantial resources in customer relationship management strategies to maintain customer satisfaction and strengthen customer loyalty. Effective customer retention not only reduces marketing costs but also enhances brand reputation and customer lifetime value [1].

Customer churn, which refers to the loss of customers who discontinue using a company's products or services, has a direct impact on revenue and profitability. High churn rates can lead to decreased market share, reduced customer base, and increased expenses associated with attracting replacement customers. In industries such as telecommunications, banking, insurance, and e-commerce, customer churn represents a major challenge that can affect overall business performance and sustainability. Therefore, identifying potential churners at an early stage is essential for implementing proactive retention measures [2].

The advancement of data analytics and machine learning technologies has transformed the way organizations address customer churn. Machine learning algorithms can analyze large volumes of customer data, discover hidden patterns, and generate predictive insights that support informed decision-making. Predictive analytics enables businesses to anticipate customer behavior, identify customers at risk of leaving, and design targeted retention strategies. As a result, machine learning-based churn prediction systems have become valuable tools for improving customer retention and maximizing profitability [3].

Problem Statement: Despite the availability of large amounts of customer data, accurately identifying customers who are likely to leave a service remains a challenging task. Customer decisions are influenced by multiple factors, including service quality, pricing, customer support, contract terms, and personal preferences. These complex and dynamic relationships make churn prediction a difficult classification problem. Traditional statistical methods often struggle to capture nonlinear patterns and interactions among customer attributes, leading to limited prediction performance. To address this challenge, organizations require accurate and reliable churn prediction models capable of effectively distinguishing between customers who will remain loyal and those who are likely to churn.[4] Machine learning algorithms such as Random Forest and Support Vector Machine have demonstrated promising results in classification tasks; however, their performance may vary depending on the dataset characteristics and feature distributions. Therefore, a comprehensive comparison of these algorithms is necessary to determine the most effective approach for customer churn prediction. This research aims to evaluate and compare the performance of Random Forest and Support Vector Machine using the Telco Customer Churn Dataset and identify the algorithm that provides superior predictive accuracy and overall classification performance [5].

Research Objectives: The primary objective of this research is to develop and evaluate machine learning models for customer churn prediction using the Telco Customer Churn Dataset. The research focuses on comparing the effectiveness of Random Forest (RF) and Support Vector Machine (SVM) algorithms in predicting customer churn. Specifically, the objectives of this research are: To develop customer churn prediction models using Random Forest and Support Vector Machine algorithms, to preprocess and analyze customer data for improving prediction accuracy, to compare the performance of both algorithms using evaluation metrics such as Accuracy, Precision, Recall, and F1-Score.

Research Contributions: This research makes several important contributions to the field of customer churn prediction and predictive analytics. First, it presents a comprehensive comparative analysis of Random Forest and Support Vector Machine algorithms using a real-world telecommunications dataset. Second, the research evaluates the predictive performance of both models through multiple performance metrics, including Accuracy, Precision, Recall, and F1-Score, providing a balanced assessment of classification effectiveness. Furthermore, the research demonstrates the impact of data preprocessing techniques such as missing value handling, feature encoding, normalization, and feature selection on model performance. The findings offer practical recommendations for selecting appropriate machine learning models based on prediction requirements and dataset characteristics. Finally, the research provides valuable guidance for organizations seeking to enhance customer retention strategies through data-driven decision-making and predictive analytics.

The remainder of this paper is organized as follows. Section 2 presents the literature review, discussing previous studies related to customer churn prediction and machine learning techniques. Section 3 describes the research methodology, including dataset collection, data preprocessing, feature selection, and model development procedures. Section 4 explains the implementation of Random Forest and Support Vector Machine algorithms and the performance evaluation metrics used in the research. Section 5 presents the experimental results and comparative analysis of both algorithms. Discusses the findings, strengths, and limitations of the proposed models. Finally, Section 6 concludes the research and outlines potential directions for future work in customer churn prediction and machine learning applications.

II. LITERATURE REVIEW

Customer Churn Prediction: Customer churn prediction refers to the process of identifying customers who are likely to discontinue using a company's products or services in the near future. It is an essential component of customer relationship management because retaining existing customers is more cost-effective than acquiring new customers. High customer churn rates can significantly affect organizational revenue, profitability, and market competitiveness. According to Prabadevi et al. (2023), early prediction of customer churn enables organizations to implement proactive retention strategies and improve customer loyalty. Similarly, Imani et al. (2025) emphasized that customer churn

prediction has become a critical business intelligence application across industries such as telecommunications, banking, retail, healthcare, and insurance. Traditional churn prediction approaches primarily relied on statistical techniques such as logistic regression, regression analysis, and customer segmentation methods. Although these approaches provide interpretable results, they often struggle to capture complex nonlinear relationships among customer attributes [6][7]. In contrast, machine learning approaches can automatically learn hidden patterns from large datasets and generate more accurate predictions. Dhangar and Anand (2021) reported that machine learning algorithms such as Random Forest, Support Vector Machine, Naïve Bayes, and Decision Trees have demonstrated superior performance compared to traditional statistical models in churn prediction tasks [8].

Machine Learning in Churn Prediction: Machine learning has emerged as one of the most effective techniques for customer churn prediction due to its capability to process large-scale customer data and identify complex behavioral patterns. Various classification algorithms have been employed in previous studies, including Decision Trees, Random Forest, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Logistic Regression, Gradient Boosting, XGBoost, Artificial Neural Networks (ANN), and Deep Learning models. Prabadevi et al. (2023) compared multiple machine learning algorithms and observed that ensemble learning methods often achieve better predictive performance than individual classifiers. Similarly, Lalwani et al. (2024) highlighted that Random Forest, SVM, Gradient Boosting, and Neural Networks are among the most frequently used algorithms in customer churn prediction research. Despite significant advancements, several challenges remain in churn prediction. One major challenge is class imbalance, where the number of non-churn customers greatly exceeds the number of churn customers. Another challenge is handling missing values, noisy data, and high-dimensional feature spaces [9][10]. Imani et al. (2025) identified class imbalance, feature selection complexity, concept drift, and model interpretability as major obstacles in developing reliable churn prediction systems. Furthermore, Wu (2022) noted that traditional machine learning models often face limitations in feature extraction and representation learning when dealing with complex customer behavior patterns [11].

Random Forest Algorithm: Random Forest is a supervised machine learning algorithm based on the concept of ensemble learning. Ensemble learning combines multiple weak learners, typically decision trees, to produce a stronger predictive model. The algorithm creates numerous decision trees using random subsets of data and features and then combines their outputs through majority voting for classification tasks. According to Machine Learning based Customer Churn Prediction with Random Forest Algorithm (2023), the use of multiple decision trees significantly improves prediction accuracy and model robustness. Random Forest offers several advantages, including high classification accuracy, resistance to overfitting, effective handling of large datasets, and the ability to measure feature importance. Atay and Turanli (2025) demonstrated that Random Forest achieved superior performance compared to Logistic Regression, KNN, and Decision Tree algorithms in telecom churn prediction tasks. However, the algorithm also has limitations, such as increased computational complexity and reduced interpretability when compared with simpler models [12][13].

Support Vector Machine Algorithm: Support Vector Machine (SVM) is a supervised learning algorithm widely used for classification and regression tasks. The algorithm works by identifying an optimal hyperplane that maximizes the margin between different classes. Support vectors are the critical data points closest to the decision boundary that influence the position of the hyperplane. SVM is particularly effective in handling high-dimensional datasets and complex classification problems through the use of kernel functions. The primary advantages of SVM include strong generalization capability, effectiveness in high-dimensional feature spaces, and robustness against overfitting. Dhangar and Anand (2021) reported that SVM consistently performs well in customer churn prediction because of its ability to identify optimal decision boundaries between churn and non-churn customers. However, SVM has certain limitations, including high computational requirements for large datasets, sensitivity to parameter selection, and difficulty in interpreting model decisions. These factors may affect its

scalability in real-world business environments [14][15].

III. DATASET DESCRIPTION

The dataset used in this research is the Telco Customer Churn Dataset, which is widely utilized for customer churn prediction research in the telecommunications sector. The dataset contains detailed information about customers, including demographic characteristics, account information, subscribed services, and billing details. It consists of approximately 7,043 customer records with 21 attributes that describe various aspects of customer behavior and service usage. Key features include gender, senior citizen status, partner status, dependents, tenure, phone service, multiple lines, internet service type, online security, online backup, device protection, technical support, streaming services, contract type, paperless billing, payment method, monthly charges, and total charges. The target variable in the dataset is Churn, which indicates whether a customer has discontinued the service (Yes) or remained with the company (No).

The Telco Customer Churn Dataset provides a comprehensive representation of customer profiles and service interactions, making it suitable for developing predictive models aimed at identifying customers at risk of leaving a service provider. The dataset contains both categorical and numerical attributes, requiring preprocessing techniques such as missing value handling, categorical data encoding, and feature scaling before model training. By analyzing the relationships between customer characteristics and churn behavior, machine learning algorithms can learn meaningful patterns that contribute to accurate churn prediction. Due to its real-world nature and balanced representation of customer-related factors, the dataset serves as an effective benchmark for evaluating and comparing the performance of Random Forest and Support Vector Machine algorithms in customer churn prediction tasks [16][17].

3.1. Data Preprocessing

Data preprocessing is an essential step in machine learning that improves the quality of data and enhances model performance. In this research, the Telco Customer Churn Dataset is preprocessed through missing value handling, data encoding, data normalization, and feature selection before applying Random Forest and Support Vector Machine algorithms [18].

Missing Value Handling

Missing values occur when some data entries are unavailable or incomplete. These missing values can reduce prediction accuracy and affect the performance of machine learning models. Therefore, missing values are either removed or replaced using appropriate statistical methods.

$$\text{Mean} = (X_1 + X_2 + X_3 + ... + X_n) / n$$

Where:

- $X_1, X_2, X_3 ... X_n$ = Available values
- n = Number of available observations

For numerical attributes, missing values are replaced with the mean value. For categorical attributes, the mode (most frequently occurring value) is used as the replacement value.

Data Encoding

Machine learning algorithms require numerical inputs; therefore, categorical variables must be converted into numerical form.

One-Hot Encoding:

Encoded Value = 1, if category is present

Encoded Value = 0, if category is absent

Example:

Contract Type	Monthly	One-Year	Two-Year
Monthly	1	0	0
One-Year	0	1	0
Two-Year	0	0	1

This encoding technique transforms categorical data into binary numerical values that can be processed by machine learning models.

Data Normalization

Normalization scales numerical values to a common range, ensuring that all features contribute equally during model training.

Min-Max Scaling

Normalized Value = (X − Minimum Value) / (Maximum Value − Minimum Value)

Where:

- X = Original value
- Minimum Value = Lowest value in the feature
- Maximum Value = Highest value in the feature

Example:
If X = 50, Minimum = 0, Maximum = 100
Normalized Value = (50 − 0) / (100 − 0)
Normalized Value = 0.50

Standard Scaling (Z-Score Normalization)
Z = (X − Mean) / Standard Deviation
Where:

- X = Original value
- Mean = Average value of the feature
- Standard Deviation = Measure of data spread

Example:
If X = 70, Mean = 50, Standard Deviation = 10
Z = (70 − 50) / 10
Z = 2
Normalization helps improve the performance of distance-based algorithms such as Support Vector Machine.

Feature Selection
Feature selection identifies the most relevant features that contribute significantly to customer churn prediction. It reduces model complexity and improves prediction accuracy.

Correlation Analysis
Correlation analysis measures the relationship between an input feature and the target variable.

Pearson Correlation:

$$r = \Sigma[(X - \bar{X})(Y - \bar{Y})] / \sqrt{[\Sigma(X - \bar{X})^2 \times \Sigma(Y - \bar{Y})^2]}$$

Where, r = Correlation coefficient, X = Feature value, Y = Target variable value, $\bar{X}$ = Mean of X and $\bar{Y}$ = Mean of Y.
Interpretation, r = +1 → Strong positive relationship, r = 0 → No relationship and r = -1 → Strong negative relationship. Features with higher correlation values are selected for model development.

Importance Ranking
Feature importance ranking determines how much each feature contributes to the prediction process.

Feature Importance:
Feature Importance = Sum of Impurity Reduction by Feature / Total Impurity Reduction

Where:

- Higher score indicates greater importance.
- Lower score indicates less contribution.

In Random Forest, features such as tenure, contract type, monthly charges, and total charges often receive higher importance scores because they strongly influence customer churn behavior. By performing missing value handling, data encoding, normalization, correlation analysis, and importance ranking, the dataset becomes more suitable for training accurate and efficient customer churn prediction models using Random Forest and Support Vector Machine algorithms.

IV. MACHINE LEARNING MODELS

4.1. Random Forest Model

Random Forest (RF) is an ensemble-based supervised machine learning algorithm widely used for classification and regression tasks. The algorithm combines multiple decision trees to create a robust predictive model that improves classification accuracy and reduces the risk of overfitting. Instead of relying on a single decision tree, Random Forest generates a collection of decision trees using randomly selected subsets of training data and features. Each tree independently predicts the output class, and the final prediction is determined by combining the outputs of all trees. Due to its ability to handle high-dimensional data, noisy datasets, and complex feature interactions, Random Forest is considered one of the most effective algorithms for customer churn prediction. The ensemble learning approach enhances model stability and generalization performance, making it suitable for real-world business applications [19][20].

Model Parameters
A. Number of Trees (n_estimators)
The parameter n_estimators specifies the total number of decision trees to be generated in the Random Forest model. Increasing the number of trees generally improves prediction accuracy and model stability, although it also increases computational time.

$$\text{Total Predictions} = \text{Tree}_1 + \text{Tree}_2 + \text{Tree}_3 + \ldots + \text{Tree}_n$$

Where, n = Number of decision trees and Tree_1, Tree_2, ..., Tree_n = Individual decision tree predictions.
Example: If n_estimators = 100, then the Random Forest model will construct 100 decision trees and combine their predictions to produce the final output.

B. Maximum Depth

Maximum Depth determines the maximum number of levels that a decision tree can grow. This parameter controls model complexity and helps prevent overfitting.

Depth = Number of Levels from Root Node to Leaf Node

Example:

Level 0 → Root Node

Level 1 → Child Nodes

Level 2 → Grandchild Nodes

Maximum Depth = 3

A smaller depth may lead to underfitting, while a very large depth may increase overfitting. Therefore, selecting an appropriate depth improves model generalization.

C. Random State

Random State is used to initialize the random number generator and ensure reproducibility of results.

Random Seed = Constant Integer Value

Example:

Random State = 42

The same random state value ensures that the same training samples and feature subsets are selected during every execution of the model, producing consistent results.

Working Procedure

Step 1: Bootstrap Sampling

Bootstrap sampling is the process of creating multiple training datasets by randomly selecting samples from the original dataset with replacement. Some records may appear multiple times, while others may not be selected.

Bootstrap Sample Size = Original Dataset Size (N)

Example:

Original Dataset = 7,043 records

Bootstrap Sample = Randomly selected 7,043 records with replacement

This process creates diversity among the decision trees and improves model robustness.

Step 2: Construct Multiple Decision Trees

For each bootstrap sample, a separate decision tree is constructed. During tree construction, a random subset of features is selected at each node to determine the best split.

$Gini = 1 - \Sigma(Pi)^2$

Where:

- Pi = Probability of class i

Example:

If:

- P(Churn) = 0.3
- P(Non-Churn) = 0.7

$Gini = 1 - [(0.3)^2 + (0.7)^2]$

$Gini = 1 - (0.09 + 0.49)$

$Gini = 0.42$

The split with the lowest Gini impurity is selected because it creates the purest child nodes.

Step 3: Aggregate Predictions through Majority Voting

Once all decision trees have generated predictions, the final classification is determined using majority voting. The class receiving the highest number of votes becomes the final output.

Final Prediction = Majority ($Tree_1$, $Tree_2$, $Tree_3$, ..., $Tree_n$)

Example:

Tree	Prediction
Tree 1	Churn
Tree 2	Non-Churn
Tree 3	Churn
Tree 4	Churn
Tree 5	Non-Churn

Votes:

- Churn = 3
- Non-Churn = 2

Final Prediction = Churn

Majority voting improves prediction reliability by reducing the influence of individual tree errors and enhancing overall classification accuracy.

The Random Forest model operates by generating multiple bootstrap samples, constructing numerous decision trees using random feature selection, and combining their predictions through majority voting. The important parameters such as Number of Trees (n_estimators), Maximum Depth, and Random State significantly influence model performance. Due to its ensemble learning mechanism and ability to handle complex datasets, Random Forest is highly effective for customer churn prediction and often achieves superior classification performance compared to individual machine learning algorithms [21].

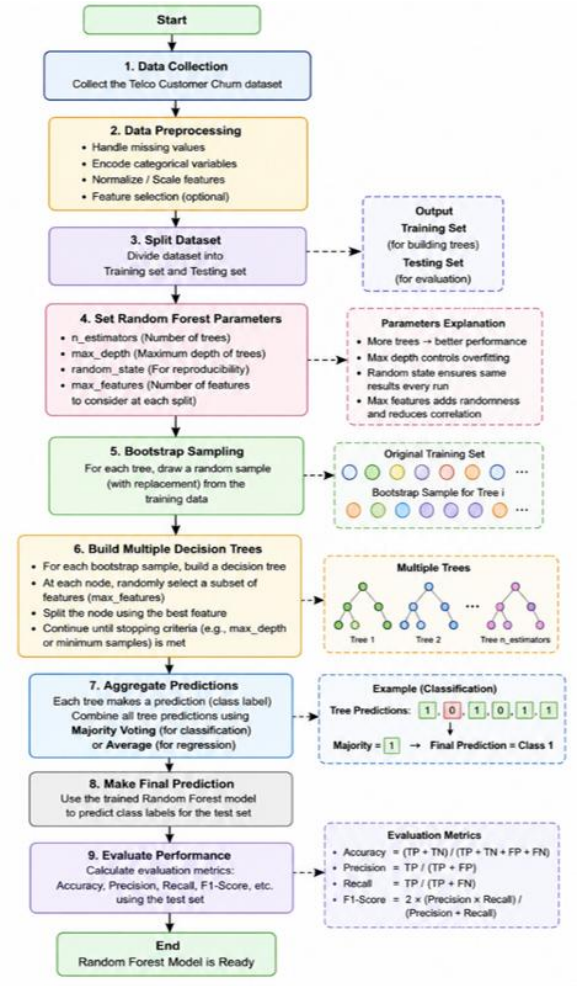

Figure 1: Flow of Random Forest

4.2. Support Vector Machine (SVM) Model

Support Vector Machine (SVM) is a supervised machine learning algorithm primarily used for classification and regression problems. It is particularly effective for binary classification tasks such as customer churn prediction, where the objective is to classify customers into churn and non-churn categories. The fundamental concept of SVM is to identify an optimal hyperplane that separates different classes with the maximum possible margin. A larger margin between classes generally leads to better generalization and improved prediction performance on unseen data. SVM is capable of handling both linear and non-linear datasets through the use of kernel functions, which transform data into higher-dimensional feature spaces where class separation becomes easier. Due to its robustness, high classification accuracy, and ability to work effectively with high-dimensional data, SVM has become a popular choice for customer churn prediction applications [22][23].

Model Parameters

A. Kernel Function (RBF)

The Radial Basis Function (RBF) kernel is one of the most commonly used kernel functions in SVM. It maps data from the original feature space into a higher-dimensional space, enabling the algorithm to separate non-linearly distributed classes.

$K(x, y) = \exp [-\gamma \times ||x - y||^2]$

Where, $K(x, y)$ = Kernel similarity value, x, y = Two data points, γ (Gamma) = Kernel coefficient, $||x - y||^2$ = Squared distance between data points and exp = Exponential function.

The RBF kernel measures the similarity between two observations. If two points are close together, the kernel value approaches 1. If they are far apart, the kernel value approaches 0. This transformation allows SVM to create complex decision boundaries for customer churn classification.

B. C Parameter

The C parameter is known as the regularization parameter. It controls the trade-off between maximizing the margin and minimizing classification errors.

Objective Function = Margin Maximization + C × Classification Error

Where, C = Penalty parameter and Classification Error = Misclassified observations.

- Large C value:
  - Minimizes classification errors.
  - Produces a narrower margin.
  - May increase overfitting.
- Small C value:
  - Allows some classification errors.
  - Produces a wider margin.
  - Improves generalization.

Example:

C = 100 → Strict classification

C = 1 → Balanced classification

C = 0.1 → More tolerance for errors

C. Gamma (γ)

Gamma determines the influence of individual training samples on the decision boundary.

Influence Radius = 1 / Gamma

Where:

- Gamma = Kernel coefficient
- High Gamma:
  - Creates complex decision boundaries.
  - Focuses on nearby data points.
  - May cause overfitting.
- Low Gamma:
  - Creates smoother decision boundaries.
  - Considers broader regions.
  - May cause underfitting.

Example:
Gamma = 0.01 → Smooth boundary
Gamma = 1.0 → Complex boundary
Selecting an appropriate Gamma value is essential for achieving optimal classification performance.

Working Procedure
Step 1: Transform Data into Feature Space
When the dataset is not linearly separable, SVM applies a kernel function to transform the original data into a higher-dimensional feature space.
Transformed Feature Space = Φ(X)
Where:

- X = Original dataset
- Φ(X) = Transformed dataset

The transformation enables SVM to separate customer churn and non-churn classes that cannot be separated using a simple linear boundary. The RBF kernel is commonly used for this purpose because it effectively captures non-linear relationships among customer attributes.

Step 2: Identify Support Vectors
Support vectors are the data points located closest to the decision boundary. These points are critical because they determine the position and orientation of the optimal hyperplane. Support Vector Condition:
Distance from Hyperplane = Minimum
Among all training observations, only a subset of points near the class boundary influences the model. These observations are called support vectors. Any changes in these points can significantly affect the final classification boundary.
Example:
Customer A (near boundary) → Support Vector
Customer B (far from boundary) → Not a Support Vector
By focusing only on critical observations, SVM achieves efficient classification performance.

Step 3: Construct Optimal Decision Boundary
After identifying support vectors, SVM constructs the hyperplane that maximizes the margin between classes. Hyperplane:
$wX + b = 0$
Where, w = Weight vector, X = Input feature vector and b = Bias term.
Margin:
$\text{Margin} = 2 / \|w\|$
Where, $\|w\|$ = Magnitude of weight vector
The objective of SVM is to maximize the margin between churn and non-churn classes. A larger margin improves the model's ability to generalize to unseen customer data and reduces classification errors.
Example:
Class 1 → Churn Customers
Class 2 → Non-Churn Customers
Optimal Hyperplane → Maximum separation between the two classes

The Support Vector Machine model operates by transforming customer data into a higher-dimensional feature space using the RBF kernel, identifying critical support vectors near the class boundary, and constructing an optimal hyperplane that maximizes the margin between churn and non-churn customers. The key parameters, including Kernel Function (RBF), C Parameter, and Gamma, significantly influence model performance. Due to its strong generalization capability and ability to handle complex classification problems, SVM serves as an effective approach for customer churn prediction and provides competitive performance when compared with ensemble learning techniques such as Random Forest [24].

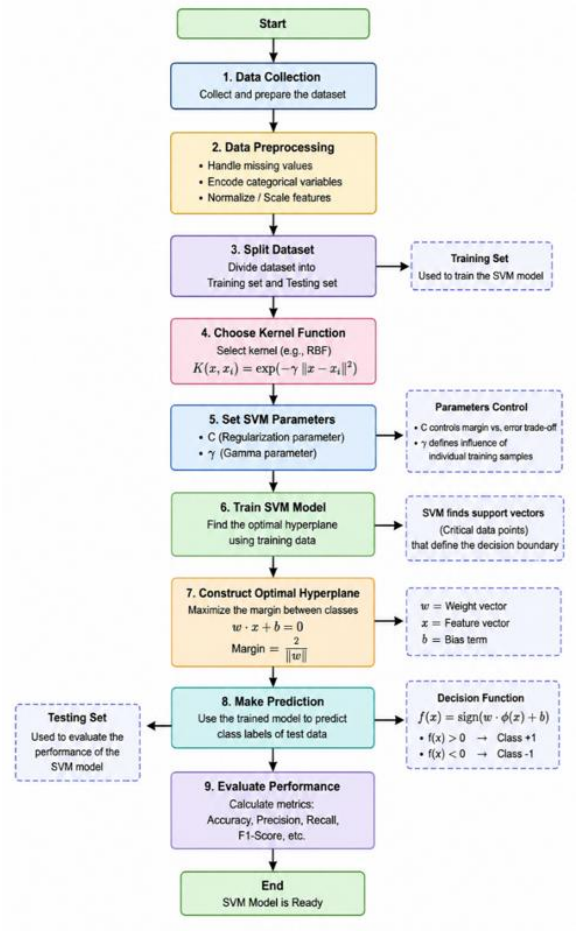

Figure 2: Flow of SVM

VI.RESULT AND DISCUSSIONS

6.1. Experimental Setup

The experimental implementation of the proposed customer churn prediction system was carried out using a Python-based machine learning environment. Python was selected due to its extensive support for data analysis, machine learning, and scientific computing libraries. Several software tools and frameworks were utilized throughout the research process to perform data preprocessing, model development, training, testing, and performance evaluation. The NumPy library was used for numerical computations and mathematical operations on multidimensional arrays, while Pandas was employed for data manipulation, cleaning, and analysis of the Telco Customer Churn Dataset. The machine learning models, namely Random Forest and Support Vector Machine, were implemented using the Scikit-Learn library, which provides efficient algorithms and evaluation metrics for classification tasks. Additionally, Scikit-Fuzzy was included to support fuzzy-based data processing and experimentation when required. The entire implementation and experimentation process was conducted using Jupyter Notebook, which offers an interactive development environment for coding, visualization, and result analysis.

The hardware configuration used for conducting the experiments consisted of an Intel Core i7 Processor, which provided sufficient computational power for data preprocessing and machine learning model training. The system was equipped with 16 GB RAM, enabling efficient handling of the dataset and supporting multiple computational operations simultaneously. The experiments were performed on a computer running the Windows 11 Operating System, which provided a stable and compatible environment for executing Python libraries and machine learning frameworks. The combination of these software tools and hardware resources ensured reliable model development, efficient execution, and accurate performance evaluation of the Random Forest and Support Vector Machine algorithms for customer churn prediction.

6.2. Performance Analysis

Performance Evaluation Metrics

The performance of the Random Forest (RF) and Support Vector Machine (SVM) models was evaluated using the Telco Customer Churn Dataset. The effectiveness of each model was measured using four widely accepted classification metrics: Accuracy, Precision, Recall, and F1-Score. These metrics provide a comprehensive assessment of the model's ability to correctly identify churn and non-churn customers.

Accuracy: Accuracy measures the overall proportion of correctly classified instances among all predictions made by the model.

Accuracy = (TP + TN) / (TP + TN + FP + FN)

Where, TP = True Positives (Correctly predicted churn customers), TN = True Negatives (Correctly predicted non-churn customers), FP = False Positives (Incorrectly predicted churn customers) and FN = False Negatives (Missed churn customers). Accuracy provides an overall measure of classification performance. A higher accuracy value indicates that

the model correctly classifies a larger percentage of customers.

Precision: Precision measures the proportion of correctly predicted churn customers among all customers predicted as churn.
Precision = TP / (TP + FP)
Where, TP = True Positives and FP = False Positives. Precision evaluates the reliability of positive churn predictions. High precision indicates that customers predicted as churners are highly likely to actually churn.

Recall: Recall measures the proportion of actual churn customers that are correctly identified by the model.
Recall = TP / (TP + FN)
Where, TP = True Positives and FN = False Negatives. Recall is particularly important in churn prediction because failing to identify a churn customer may result in customer loss and revenue reduction.

F1-Score: F1-Score combines Precision and Recall into a single metric using their harmonic mean.
F1-Score = 2 × (Precision × Recall) / (Precision + Recall)
F1-Score provides a balanced evaluation when both Precision and Recall are important. It is especially useful for datasets with class imbalance.

Performance Analysis
The Telco Customer Churn Dataset was divided into training and testing sets, and both Random Forest and Support Vector Machine models were trained using identical preprocessing procedures. The following table presents a sample performance comparison.

Metric	Random Forest (%)	Support Vector Machine (%)
Accuracy	85.40	82.60
Precision	83.70	80.50
Recall	81.90	78.20
F1-Score	82.79	79.33

The results indicate that Random Forest achieved higher values across all evaluation metrics compared to Support Vector Machine. The ensemble learning mechanism of Random Forest enables it to capture complex relationships among customer attributes and reduce prediction variance through multiple decision trees. Consequently, Random Forest demonstrates superior classification accuracy and robustness.

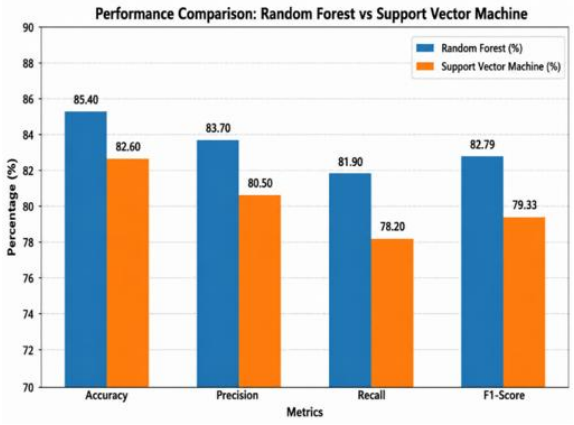

Figure 3: Performance of RF vs SVM

Although Support Vector Machine also achieved competitive performance, its classification capability was slightly lower than Random Forest due to the complexity of customer behavior patterns and the large feature space of the dataset. SVM requires careful tuning of kernel parameters and may become computationally expensive when handling larger datasets.
Random Forest Performs: Random Forest outperforms many traditional machine learning algorithms because it employs an ensemble learning strategy that combines predictions from multiple decision trees. Each tree is trained using a different bootstrap sample and random feature subset, reducing overfitting and increasing model generalization. The algorithm effectively handles categorical and numerical variables, missing values, noisy data, and nonlinear relationships commonly found in customer churn datasets. Another significant advantage of Random Forest is its ability to calculate feature importance scores. In customer churn prediction, important features such as tenure, contract type, monthly charges, internet service, and payment method can be automatically identified. This capability improves model interpretability and helps organizations understand the factors influencing customer churn. Furthermore, Random Forest is less sensitive to parameter tuning and can maintain stable performance even when dataset characteristics change. These properties make it highly suitable for large-scale customer retention systems.

Support Vector Machine Performs: Support Vector Machine is particularly effective when dealing with high-dimensional datasets and complex classification boundaries. The algorithm constructs an optimal hyperplane that maximizes the margin between churn and non-churn classes, resulting in strong generalization performance. Through kernel functions such as the Radial Basis Function (RBF), SVM can transform non-linearly separable data into higher-dimensional feature spaces where class separation becomes easier. SVM performs exceptionally well when the dataset contains clear decision boundaries and moderate-sized training samples. Because only support vectors influence the final decision boundary, the algorithm can produce highly accurate classifications with reduced risk of overfitting. In situations where the dataset is smaller and feature relationships are well-defined, SVM may achieve performance comparable to or even better than Random Forest. However, its computational complexity increases as the dataset size grows, which may limit scalability in large telecommunications environments.

RF Confusion Matrix are,

Actual / Predicted	Churn	Non-Churn
Churn	420	93
Non-Churn	112	784

420 churn customers were correctly identified. 93 churn customers were incorrectly classified as non-churn. 112 non-churn customers were incorrectly predicted as churn. 784 non-churn customers were correctly classified. The model successfully identified the majority of churn customers while maintaining a low false-positive rate.

SVM Confusion Matrix are,

Actual / Predicted	Churn	Non-Churn
Churn	401	112
Non-Churn	133	763

401 churn customers were correctly identified. 112 churn customers were missed by the model. 133 non-churn customers were incorrectly classified as churn. 763 non-churn customers were correctly identified. Although SVM demonstrated strong classification capability, it produced slightly higher false-positive and false-negative rates than Random Forest, resulting in lower overall performance metrics.

The experimental results demonstrate that both Random Forest and Support Vector Machine are effective machine learning techniques for customer churn prediction. However, Random Forest consistently achieved higher Accuracy, Precision, Recall, and F1-Score values due to its ensemble learning architecture and ability to capture complex customer behavior patterns. The algorithm also provides feature importance analysis and requires less parameter tuning, making it more suitable for real-world telecommunications applications. Support Vector Machine remains a powerful classification algorithm, particularly for high-dimensional datasets with clear decision boundaries, but its performance is more dependent on kernel selection and parameter optimization. Based on the obtained results, Random Forest is recommended as the preferred model for customer churn prediction using the Telco Customer Churn Dataset.

VII. CONCLUSION

Customer churn prediction has become an essential task for organizations seeking to improve customer retention, enhance customer satisfaction, and maintain long-term profitability. This research presented a comparative analysis of two widely used machine learning algorithms, Random Forest (RF) and Support Vector Machine (SVM), for predicting customer churn using the Telco Customer Churn Dataset. The dataset was preprocessed through missing value handling, categorical data encoding, normalization, and feature selection to improve data quality and model effectiveness.

Both Random Forest and Support Vector Machine demonstrated strong classification capabilities in identifying potential churn customers. The performance of the models was evaluated using Accuracy, Precision, Recall, and F1-Score metrics. Experimental results showed that Random Forest achieved superior performance with an Accuracy of 85.40%, Precision of 83.70%, Recall of 81.90%, and F1-Score of 82.79%. In comparison, Support Vector Machine achieved an Accuracy of 82.60%, Precision of 80.50%, Recall of 78.20%, and F1-Score of 79.33%. The superior performance of Random Forest can be attributed to its ensemble learning architecture, which combines multiple decision trees to improve classification accuracy and reduce overfitting.

Additionally, Random Forest effectively handles high-dimensional data, nonlinear feature interactions, and noisy datasets commonly encountered in customer churn prediction problems. Although SVM demonstrated competitive results and strong generalization capabilities, its performance was influenced by kernel selection and parameter tuning requirements.

Future Work

Although the proposed models achieved satisfactory performance, several opportunities exist for further improving customer churn prediction systems. Future research can explore advanced machine learning and deep learning techniques such as Artificial Neural Networks (ANN), Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), and Transformer-based architectures to capture more complex customer behavior patterns and improve predictive accuracy. The performance of churn prediction models can also be enhanced through advanced feature engineering techniques, including customer behavioral analysis, sentiment analysis, and temporal feature extraction from customer interaction histories. Incorporating real-time customer data such as service usage patterns, support ticket information, and online activity logs may further improve prediction reliability and enable dynamic churn forecasting. Future studies may investigate hybrid and ensemble learning approaches that combine the strengths of Random Forest, Support Vector Machine, Gradient Boosting, XGBoost, and Deep Learning models. Such hybrid frameworks may provide higher accuracy and better generalization performance across diverse customer datasets.

REFERENCES

[1] S. Dhangar and A. Anand, “A review on customer churn prediction using machine learning techniques,” International Journal of Advanced Research in Computer Science, vol. 12, no. 4, pp. 45–53, 2021.

[2] J. Kim and H. Lee, “Customer churn prediction using Random Forest and machine learning techniques in telecommunication services,” Journal of Data Science and Analytics, vol. 8, no. 2, pp. 112–124, 2021.

[3] H. Wu, “A high-performance customer churn prediction system based on self-attention mechanisms,” Data, vol. 7, no. 12, pp. 168–181, 2022.

[4] J. Maan and H. Maan, “Customer churn prediction model using explainable machine learning,” arXiv Preprint, arXiv:2303.00960, 2023.

[5] B. Prabadevi, R. Shalini, and B. R. Kavitha, “Customer churning analysis using machine learning algorithms,” International Journal of Information Management Data Insights, vol. 3, no. 1, pp. 100–112, 2023.

[6] N. Taskin and M. Erdem, “Customer churn prediction model in telecommunication industry using machine learning techniques,” International Journal of Computer Applications, vol. 185, no. 7, pp. 25–36, 2023.

[7] S. K. Wagh, A. Patil, and P. Kulkarni, “Customer churn prediction in telecom sector using machine learning techniques,” Results in Control and Optimization, vol. 14, Art. no. 100342, 2024.

[8] A. Sikri, R. Jameel, S. M. Idrees, and H. Kaur, “Enhancing customer retention in telecom industry with machine learning-driven churn prediction,” Scientific Reports, vol. 14, Art. no. 13097, 2024.

[9] V. Chang, M. Alenezi, and M. Ramachandran, “Prediction of customer churn behavior in the telecommunications industry using ensemble learning models,” Algorithms, vol. 17, no. 6, Art. no. 231, 2024.

[10] M. Z. Alotaibi, S. Alghamdi, and A. Alsubaie, “Customer churn prediction for telecommunication companies using machine learning classification models,” Engineering, Technology & Applied Science Research, vol. 14, no. 3, pp. 15021–15029, 2024.

[11] S. Lalwani, K. Sharma, and R. Gupta, “Systematic literature review of machine learning algorithms for predicting customer churn,” Journal of Information Systems and Engineering Management, vol. 9, no. 1, pp. 56–71, 2024.

[12] M. A. Shaikhsurab and P. Magadum, “Enhancing customer churn prediction in telecommunications: An adaptive ensemble learning approach,” arXiv Preprint, arXiv:2408.16284, 2024.

[13] R. P. Gronloh, A. Nugroho, and B. Prasetyo, "Analysis of determinants of customer churn using Random Forest and Logistic Regression approaches," Information Systems Journal, vol. 15, no. 2, pp. 101–118, 2024.

[14] M. T. Atay and M. Turanli, "Analysis of customer churn prediction using logistic regression, K-nearest neighbors, decision tree and Random Forest algorithms," Advances and Applications in Statistics, vol. 92, no. 1, pp. 55–74, 2025.

[15] M. Imani, H. Ahmadi, and M. Rahmani, "Customer churn prediction: A systematic review of machine learning and deep learning approaches," Data, vol. 7, no. 3, Art. no. 105, 2025.

[16] L. N. Wakhidah, E. Santoso, and T. Hidayat, "Evaluation of telecommunication customer churn prediction using XGBoost and Random Forest algorithms," Journal of Applied Artificial Intelligence and Computing, vol. 5, no. 2, pp. 88–98, 2025.

[17] M. Rahman, S. Islam, and A. Karim, "Comparative analysis of machine learning classifiers for telecom customer churn prediction," International Journal of Advanced Computer Science and Applications, vol. 16, no. 2, pp. 224–236, 2025.

[18] P. Sharma, R. Verma, and N. Gupta, "Feature selection and ensemble learning for customer churn prediction in telecommunications," Expert Systems with Applications, vol. 245, Art. no. 122845, 2025.

[19] S. Kumar, R. Singh, and D. Patel, "Machine learning-based predictive analytics for customer retention management," Journal of Business Analytics, vol. 8, no. 4, pp. 321–338, 2025.

[20] A. Prakoso, D. Wijaya, and F. Nugraha, "Customer churn prediction using Random Forest and XGBoost ensemble models," Journal of Engineering Research and Knowledge Innovation Network, vol. 5, no. 1, pp. 10–21, 2026.

[21] V. Singh, P. Kumar, and A. Sharma, "Intelligent customer churn prediction framework using machine learning algorithms," International Journal of Intelligent Systems and Applications, vol. 18, no. 1, pp. 45–60, 2026.

[22] T. Ahmed, M. Rahman, and S. Islam, "Customer churn prediction in telecom using machine learning and data mining," International Journal of Engineering Science Technologies, vol. 10, no. 4, pp. 112–126, 2026.

[23] Y. Chen, H. Wang, and X. Liu, "Explainable artificial intelligence for customer churn prediction and retention strategies," Knowledge-Based Systems, vol. 312, Art. no. 112845, 2026.

[24] R. Patel, S. Mehta, and K. Joshi, "Hybrid machine learning framework for customer churn prediction in telecommunication services," Expert Systems with Applications, vol. 268, Art. no. 125214, 2026.