

PROJECT TITLE: CUSTOMER CHURN PREDICTION

PRESENTED BY

STUDENT NAME: RAJAT

COLLEGE NAME:M.B.M UNIVERSITY

**DEPARTMENT:INFORMATION
TECHNOLOGY**

**EMAIL
ID:RAJATTSHARMA87077@GMAIL.COM**

AICTE STUDENT ID:AINSI_125031



OUTLINE

- **Problem Statement**
- **Proposed System/Solution**
- **System Development Approach (Technology Used)**
- **Algorithm & Deployment**
- **Result (Output Image)**
- **Conclusion**
- **Future Scope**
- **References**

PROBLEM STATEMENT

Customer churn is a critical issue for businesses, as losing customers directly impacts revenue and growth. Predicting churn helps companies take proactive measures to retain customers. The challenge lies in accurately identifying customers who are likely to churn based on historical data, demographics, usage patterns, and other relevant factors.

PROPOSED SOLUTION

- The proposed system aims to predict customer churn using a **Random Forest Classifier**, a robust machine learning algorithm suitable for classification tasks. The solution involves:
- **Data Collection:**
 - Gather historical customer data, including features like contract type, usage frequency, payment history, and customer service interactions.
- **Data Preprocessing:**
 - Handle missing values, outliers, and categorical variables (e.g., one-hot encoding).
 - Perform feature scaling and selection to improve model performance.
- **Machine Learning Algorithm:**
 - Implement a **Random Forest Classifier** due to its ability to handle high-dimensional data and reduce overfitting.
 - Train the model on labeled data (churned vs. retained customers).
- **Deployment:**
 - Develop a user-friendly dashboard or API to provide real-time churn predictions.
 - Deploy the model using platforms like Flask, Django, or cloud services (e.g., AWS, Heroku).

SYSTEM APPROACH

1. System Requirements:

- Python 3.x, Jupyter Notebook, scikit-learn, Pandas, NumPy, Matplotlib/Seaborn.
- Libraries: sklearn.ensemble.RandomForestClassifier, imbalanced-learn (if data is skewed).

2. Technology Stack:

- Backend: Flask/Django for deployment.
- Frontend: HTML/CSS/JavaScript (optional for dashboard).

ALGORITHM & DEPLOYMENT

- **Algorithm Selection:**

- Random Forest is chosen for its ensemble learning approach, which improves accuracy by aggregating multiple decision trees.

Data input:

Features: Tenure, Monthly charges, contract Type, Internet Services, etc

Target Variable: Churn (Binary: Yes/No).

Training Process:

Split data into training (70%) and testing (30%) sets.

Use GridSearchCV for hyperparameter tuning.

Prediction Process:

The trained model outputs churn probabilities; a threshold(e.g. 0.5) classifies costumers as churn /not churn.

RESULT

- **Output:**
 - Confusion matrix, ROC curve, and feature importance plot (see below).
 - Example: **Accuracy: 85%, Precision: 82%, Recall: 78%.**

CONCLUSION

The Random Forest model effectively predicts customer churn with high accuracy. Key challenges included handling imbalanced data and selecting optimal features. The solution empowers businesses to reduce churn by targeting at-risk customers with retention strategies.

FUTURE SCOPE

- Incorporate **real-time data** (e.g., customer feedback, social media sentiment).
- Experiment with **deep learning models** (e.g., Neural Networks) for complex patterns.
- Expand to **multi-industry applications** (e.g., telecom, banking).
-

REFERENCES

- Breiman, L. (2001). *Random Forests*. Machine Learning, 45(1), 5-32.
- Scikit-learn Documentation: <https://scikit-learn.org>

.

GitHub Link: <https://github.com/Rajat-Shrma/customerchurnproject>

Thank you

