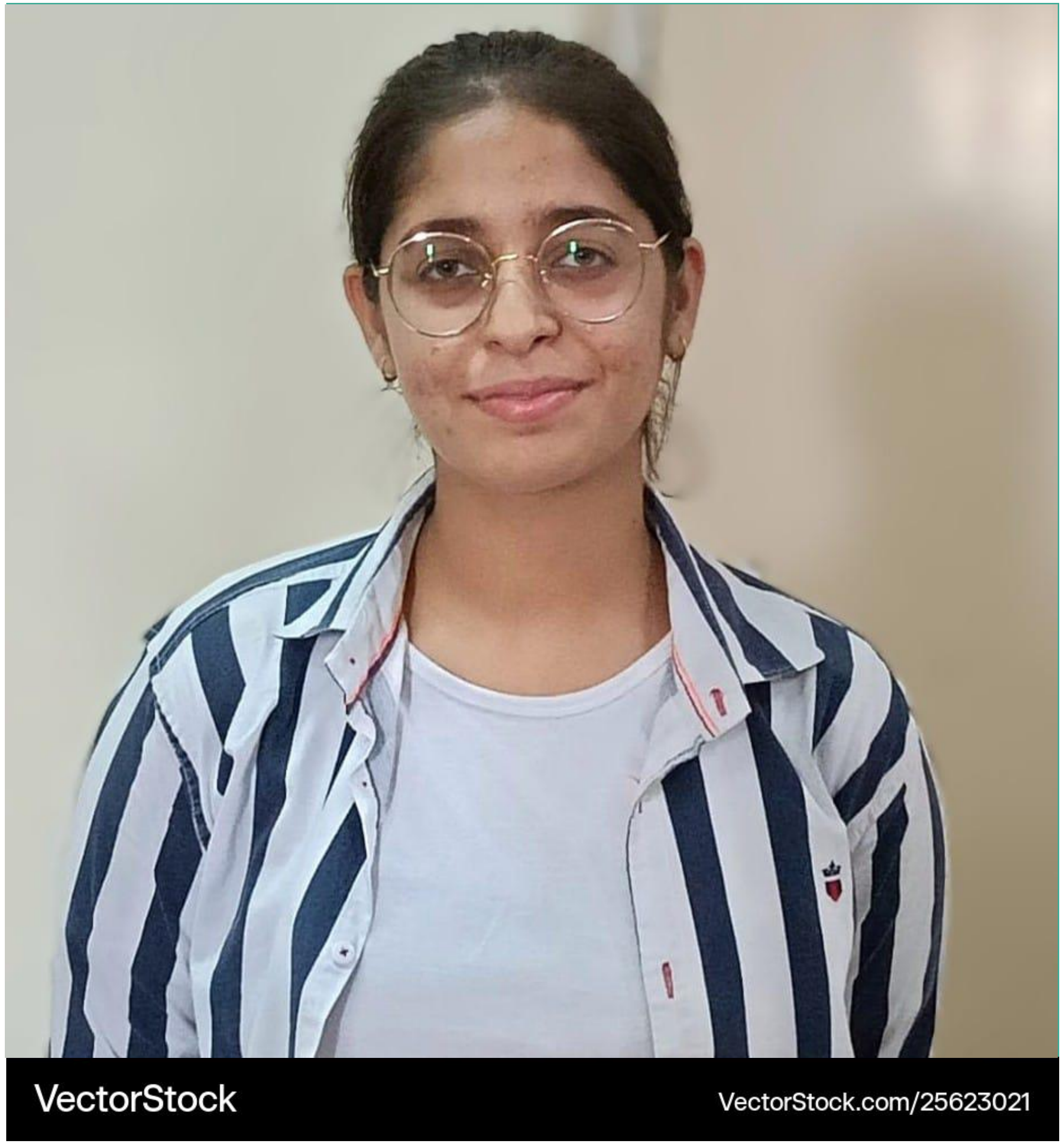
CAPSTONE PROJECT

Customer Churn Prediction Using AI/ML



PRESENTED BY

STUDENT NAME : ANSHU

COLLEGE NAME : CH.DEVI LAL STATE INSTITUTE OF ENG & TECH.

DEPARTMENT : CSE

EMAIL ID : ANSHUAR202@GMAIL.COM

AICTE STUDENT ID:STU68146F33C3EDE1746169651

OUTLINE

- **Problem Statement** (Should not include solution)
- **Proposed System/Solution**
- **System Development Approach** (Technology Used)
- **Algorithm & Deployment**
- **Result (Output Image)**
- **Conclusion**
- **Future Scope**
- **References**

PROBLEM STATEMENT

- In the highly competitive telecom industry, retaining existing customers is crucial for sustained business growth. However, customer churn—the phenomenon where customers discontinue their services—remains a significant challenge. Understanding and predicting churn is essential to enable timely interventions and reduce revenue loss.
- Currently, telecom companies lack effective systems to proactively identify customers who are likely to leave. The challenge lies in analyzing large volumes of customer data and identifying key factors influencing churn behavior.
- This project aims to address this issue by building a machine learning model that predicts customer churn based on historical and behavioral data, enabling companies to take preventive actions and improve customer retention.

PROPOSED SOLUTION

- **Data Collection:**

  Gather historical customer data including demographics, services used, billing information, and churn status from the Telco Customer Churn dataset.

- **Data Preprocessing:**
  - Handle missing values
  - Convert categorical variables
  - Perform feature selection
  - Apply scaling/normalization (if required)

- **Data Balancing:**

  Use **SMOTE (Synthetic Minority Oversampling Technique)** to handle class imbalance between churned and non-churned customers.

- **Model Training:**

  Train multiple machine learning models including:
  - Decision Tree
  - Random Forest
  - XGBoost

- **Model Evaluation:** Evaluate each model using metrics like:
  - Accuracy
  - Precision, Recall, F1-score
  - Confusion Matrix
  - ROC-AUC Curve

- **Result Analysis:** Identify the best-performing model (Random Forest) and extract feature importances to interpret churn behavior.

SYSTEM APPROACH

- The system is developed using a structured machine learning workflow:
- **Data Collection**
  - Source: Telco Customer Churn dataset (Kaggle)
  - Attributes: Customer demographics, services subscribed, billing information, churn label
- **Data Preprocessing**
  - Handling missing values
  - Encoding categorical variables (Label/One-Hot Encoding)
  - Removing duplicates and irrelevant features
- **Data Balancing**
  - Applied **SMOTE (Synthetic Minority Oversampling Technique)** to resolve class imbalance
- **Model Building**
  - Implemented and trained models:
    - Decision Tree
    - Random Forest
    - XGBoost
- **Model Evaluation**
  - Used evaluation metrics: Accuracy, Precision, Recall, F1-score, ROC-AUC
  - Performed 5-fold cross-validation for robustness
- **Result Analysis**
  - Random Forest gave the highest accuracy (~80%)
  - Feature importance showed key churn indicators like contract type, monthly charges, and tenure

ALGORITHM & DEPLOYMENT

- **Algorithm Selection:-**

We selected **Random Forest Classifier** as the primary algorithm due to its robustness, high accuracy, and ability to handle both numerical and categorical features. We also compared it with Decision Tree and XGBoost to identify the most effective model.

- **Data Input:-**

The input features used for training the model include:

Customer tenure

Monthly and total charges

Internet service type

Contract type (e.g., month-to-month, one year, two years)

Payment method

Senior citizen status

Number of dependents

And many more categorical and numerical variables

- **Training Process:-**

The dataset was split into training and testing sets (80-20 ratio).

Applied **SMOTE** to balance the classes before training.

Performed **5-fold cross-validation** for robust evaluation.
Hyperparameters were tuned using GridSearchCV (where applicable).

❑ Prediction Process:

•Evaluation metrics like Accuracy, Precision, Recall, F1-score, and ROC-AUC were used.
•The trained model predicts whether a customer is likely to churn (Yes/No) based on their historical and current profile.
•The model is currently run in an offline environment (Jupyter Notebook) without real-time deployment.

RESULT

- **Model Performance Summary (Random Forest):**
  Accuracy: 77.57% (~78%)
  Precision (Class 0): 0.85 **(Churn = No)**
  Precision (Class 1): 0.58 **(Churn = Yes)**
  F1-score: 0.78 *(weighted average)*
  ROC-AUC: 0.83 *(from earlier test)*

- **Confusion Matrix:**
  [[877 159] → Actual: Not Churned [157 216]] → Actual: Churned

- **Interpretation:**
  Model performs very well for **non-churned customers**
  SMOTE helped improve performance on **minority class (churned)**, **but slight imbalance still remains**

```
Accuracy Score:
 0.7757274662881476
Confusion Matrix:
 [[877 159]
 [157 216]]
Classification Report:
              precision    recall  f1-score   support

           0       0.85      0.85      0.85      1036
           1       0.58      0.58      0.58       373

    accuracy                           0.78      1409
   macro avg       0.71      0.71      0.71      1409
weighted avg       0.78      0.78      0.78      1409
```

Figure 1: Classification Report & Confusion Matrix Output

CONCLUSION

- Successfully developed a machine learning model to predict customer churn using the Telco dataset.
- Applied data preprocessing, SMOTE for class imbalance, and implemented multiple ML algorithms (Decision Tree, Random Forest, XGBoost).
- Among all models, **Random Forest** provided the best performance with ~78% accuracy and good generalization.

❑ **Key Takeaways**

- **Contract type, tenure, and monthly charges** were identified as major factors influencing customer churn.
- **SMOTE** improved detection of the minority class (churned customers), leading to better balance and fairness.
- The model can help telecom companies **proactively target at-risk customers**, thereby improving retention and reducing revenue loss.

❑ **Impact**

- The project demonstrates the real-world business value of AI/ML in customer behavior analysis.
- A deployable predictive system could support smarter decision-making in customer relationship management.

FUTURE SCOPE

- **Model Deployment**:

  Deploy as a web or cloud-based churn prediction service.
- **Advanced Algorithms**:

  Explore deep learning or ensemble stacking for improved accuracy.
- **More Data Sources**:

  Integrate customer interaction data like support calls, usage logs, or complaint history.
- **Real-Time Prediction**:

  Enable real-time churn alerts using APIs or streaming data.
- **Scalability**:

  Extend the model to other industries or service-based companies facing churn issues.

REFERENCES

•**Telco Customer Churn Dataset (Kaggle):**
https://www.kaggle.com/blastchar/telco-customer-churn
•**scikit-learn: Machine Learning Library**
https://scikit-learn.org/stable/
•**XGBoost: Gradient Boosting Algorithm**
https://xgboost.readthedocs.io/
•**imbalanced-learn (SMOTE Technique):**
https://imbalanced-learn.org/stable/over_sampling.html#smote
•**GitHub Link** (Project Code):
https://github.com/Anshu-code-202/MSLearn-Customer-Churn-Prediction

Thank you