Bangladesh Army University of Science and Technology (BAUST) Saidpur Cantonment, Nilphamari

Department of Computer Science and Engineering (CSE)

Course Code: CSE 4140

Course Title: Machine Learning Sessional

Customer Churn Prediction using Machine Learning

Presented By

Sumiya Naznin Haque
ID: 220201020
Md Abdullah
ID: 220201034

Presented To

Engr. Rohul Amin
Lecturer, CSE
Nadim Reza
Lecturer, CSE

Table of contents

01 Introduction

02 Objectives

03 Related Works

04 Dataset

05 Methodology

06 Models Used

07 Comparison Table

08 Our Improvements

09 Conclusion

10 References

Introduction

- Customer churn refers to the loss of existing customers over a given period.
- Telecom industries face significant revenue loss due to high churn rates.
- Machine Learning enables data-driven identification of customers likely to churn.

Key Focus

- Identify churn vs. non-churn customers
- Support decision-making using data-driven insights

Objectives

- To detect customers who are likely to churn.
- To analyze key features that influence churn.
- To apply machine learning algorithms for churn prediction.
- To evaluate model performance using standard ML metrics.

Related Works

In [1], Ahmad et al., 2019 proposed a “Customer Churn Prediction Framework Using Machine Learning in Big Data Platform.”

Shortcomings:

•Features were not easy to interpret.

•Did not evaluate using ROC curves or thresholds.

•Did not fully handle class imbalance in the data.

Related Works (Cont.)

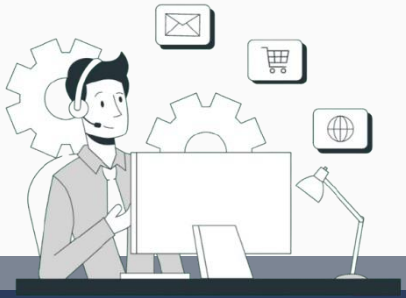
In [2], Vafeiadis et al., 2015
conducted “A Comparison of Machine Learning Techniques for Customer Churn Prediction.”
Shortcomings:

- Focused mainly on accuracy, ignoring other important metrics.
- Did not analyze performance per class (no confusion matrix).
- Limited handling of noisy or imbalanced data.

Dataset

[3] Kaggle Customer Churn Dataset

- 7043 customers
- Features: Tenure, Monthly Charges, Internet Service, Contract Type, etc.
- Target Variable: Churn (Yes/No)
- Highly imbalanced dataset (more Non-Churn than Churn)



Methodology

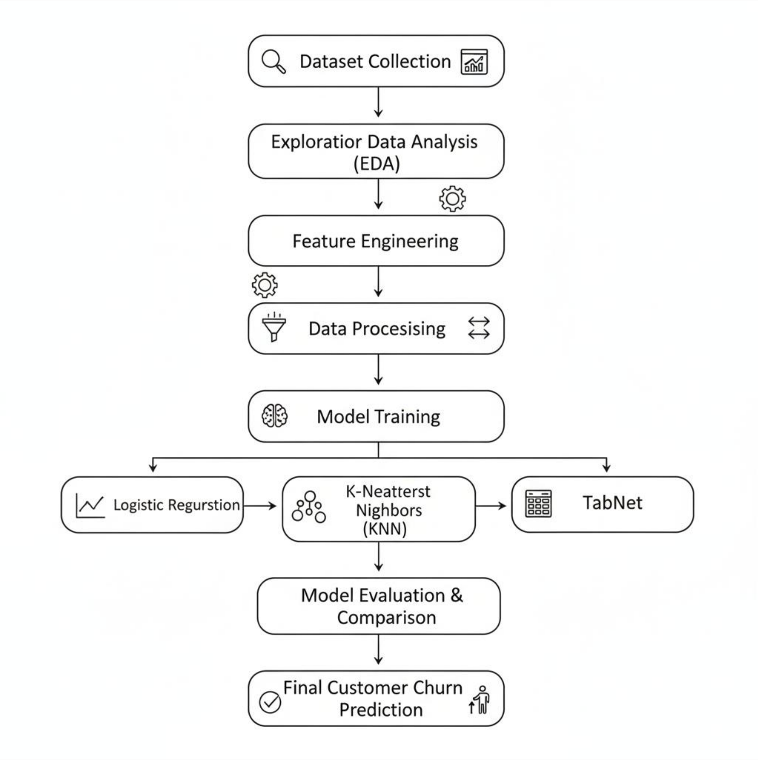


Figure-01: Methodology

Models Used

- Logistic Regression
- K-Nearest Neighbors
- Decision Tree (Tebnet)

Models Used (Cont.)

Logistic Regression Results: Accuracy: 79%

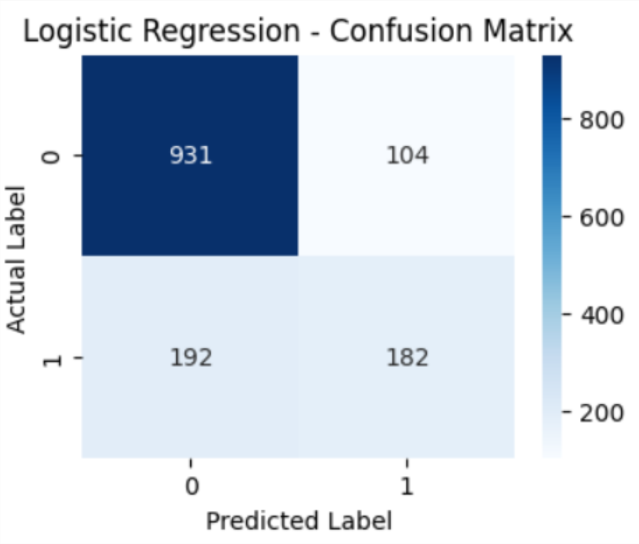


Figure-02: Confusion matrix for Logistic Regression Model

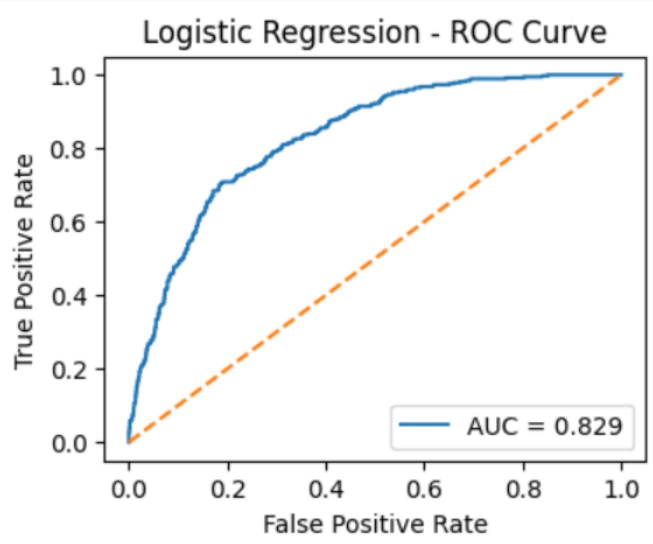


Figure-03: ROC Curve for Logistic Regression Model

Model Used (Cont.)

KNN Model Results: Accuracy: 78%

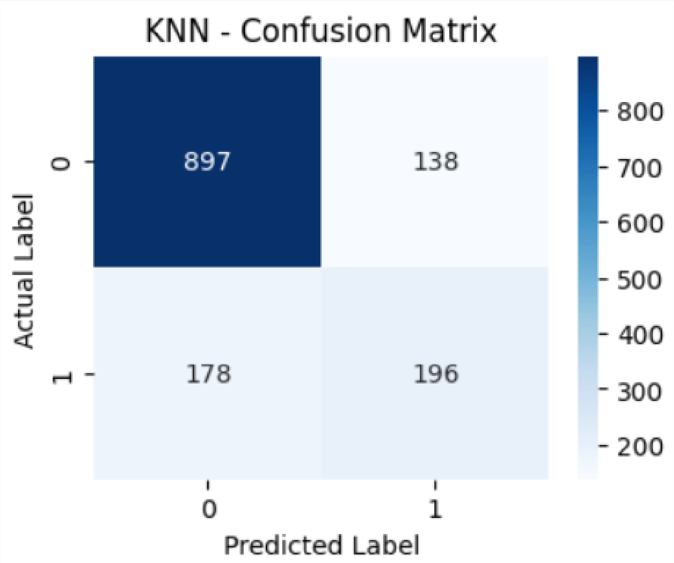


Figure-04: Confusion matrix for KNN Model

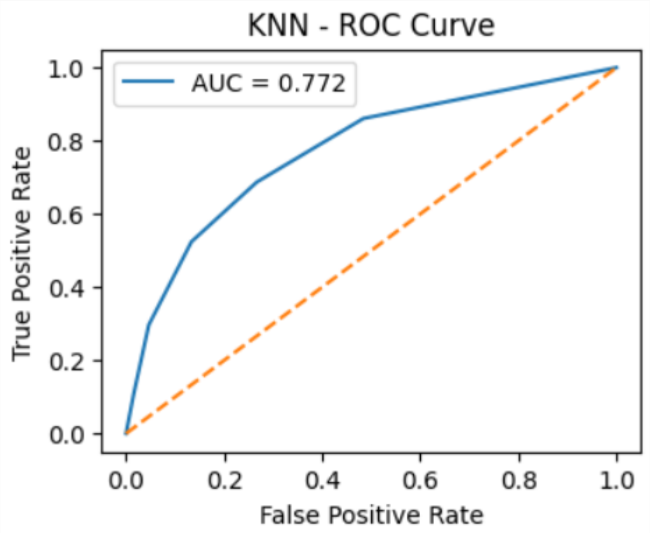


Figure-05: ROC Curve for KNN Model

Model Used (Cont.)

Decision Tree (Tebnet) Results: Accuracy: 72%

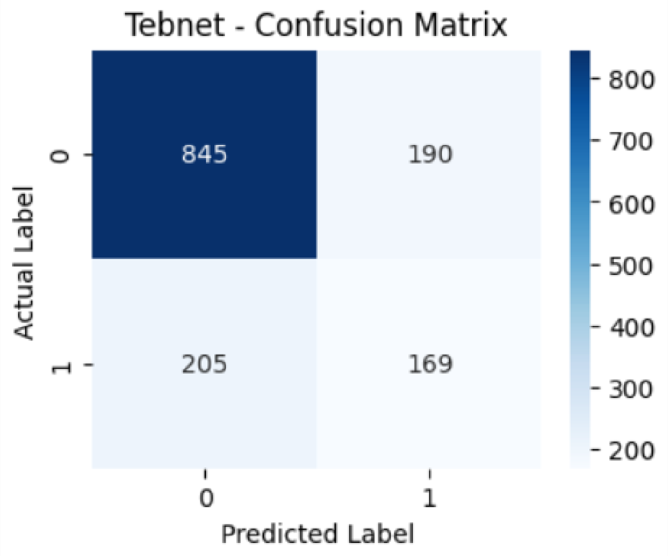


Figure-06: Confusion matrix for Tebnet Model

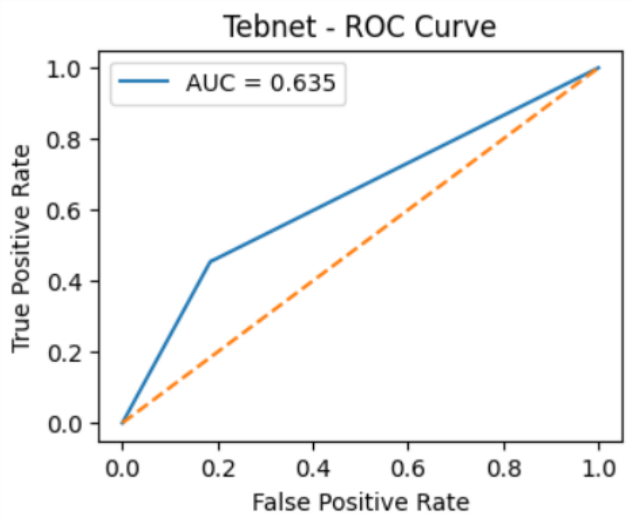


Figure-07: ROC Curve for Tebnet Model

Comparison Table

Model	Precision	Recall	F1-Score	Accuracy
Logistic Regression	0.78	0.79	0.78	0.79
KNN	0.77	0.78	0.77	0.78
Decision Tree	0.72	0.72	0.72	0.72

Our Improvements

Paper	Shortcomings	How Our Model Improves
Ahmad et al., 2019	No ROC	Includes ROC curves
Vafeiadis et al., 2015	Focused only on accuracy; No class-level evaluation	Uses multiple metrics (precision, recall, F1); Confusion matrix for per-class performance

Conclusion

- Logistic Regression showed best performance.
- KNN provided balanced results but still struggled with churn class.
- Decision Tree overfitted and performed worst.

References

[1] A. K. Ahmad, A. Jafar, and K. Aljoumaa, *Journal of Big Data*, vol. 6, no. 28, 2019.

[2] S. Vafeiadis et al., *International Journal of Computer Applications*, vol. 127, no. 8, 2015.

[3] Kaggle, "Customer Churn Dataset," Kaggle, 2018. [Online]. Available: https://www.kaggle.com/datasets/rashadrmammadov/customer-churn-dataset.

Thanks

Any Questions?