Enhancing Explainability in Machine Learning Models for Customer Churn Prediction

Presented by:

Name: Sai Kumari Lukka

Student ID: S3308548

Introduction

- Customer churn is a major challenge in the telecom industry, impacting revenue and growth.
- This project analyzes and preprocesses telecom customer data to identify churn-related patterns.
- Machine learning classification models are used to predict customer churn accurately.
- Model performance is evaluated using standard metrics to ensure reliability.
- The final model is deployed as a web-based application for real-time predictions.

Importance of the Study

- Customer churn significantly impacts revenue in the telecom industry, making accurate prediction essential (Verbeke et al., 2012).
- Many machine learning models act as black boxes, limiting trust and business usability (Molnar, 2022).
- Explainable AI techniques such as SHAP and LIME improve transparency and interpretability of churn predictions (Lundberg and Lee, 2017; Ribeiro et al., 2016).
- Descriptive analytics provides deeper insight into customer behavior, supporting better retention decisions (Kohavi, 1995).

Literature Review

Key Studies:

- Verbeke et al. (2012): Machine learning models are effective for telecom churn prediction.
- Molnar (2022): Highlights the lack of transparency in black-box machine learning models and the need for explainability.
- Lundberg and Lee (2017): SHAP provides consistent and interpretable explanations for complex models.

Research Focus:

- Focuses on explainability rather than only predictive accuracy.
- Integrates descriptive analytics for business understanding.
- Supports transparent and ethical churn prediction.

Research Question and Objectives

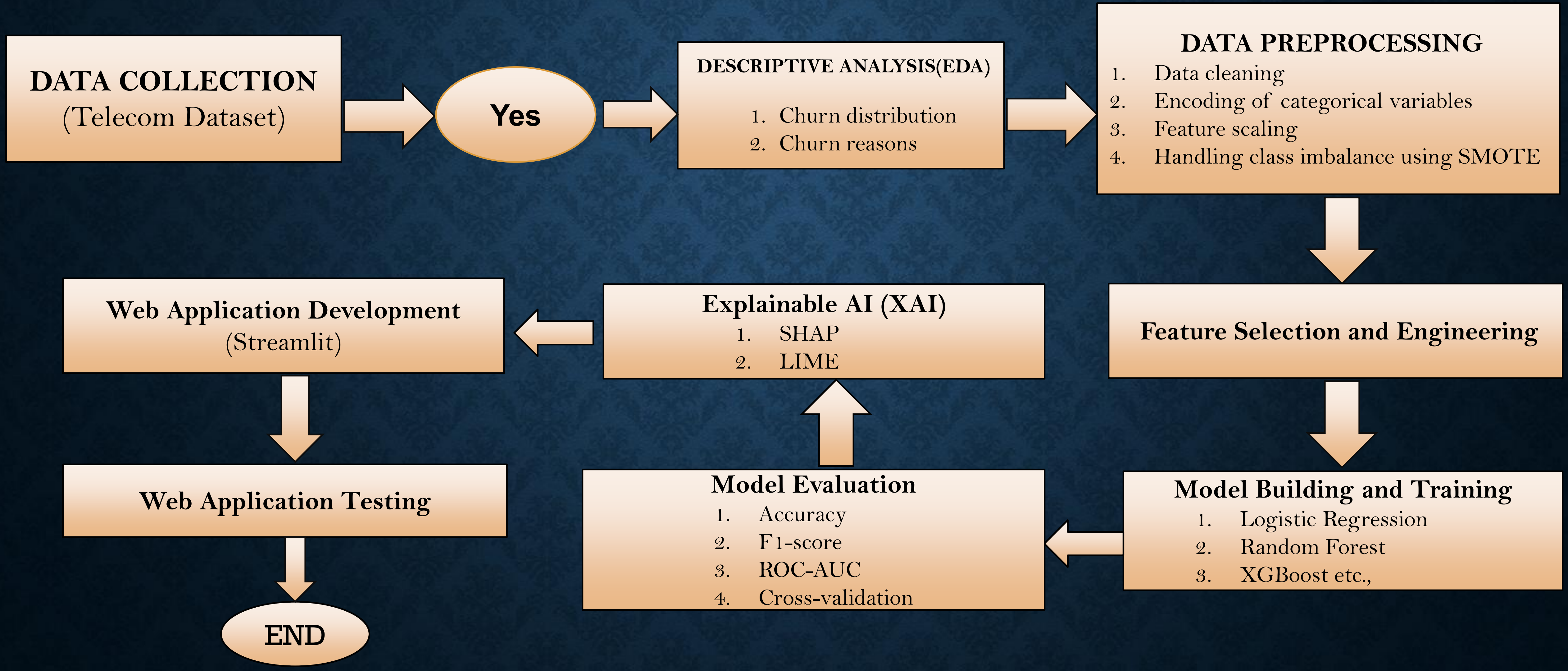
Research Question:

How can an explainable machine learning framework be designed to accurately predict customer churn and identify the underlying reasons for churn in the telecommunications industry?

Research Objectives:

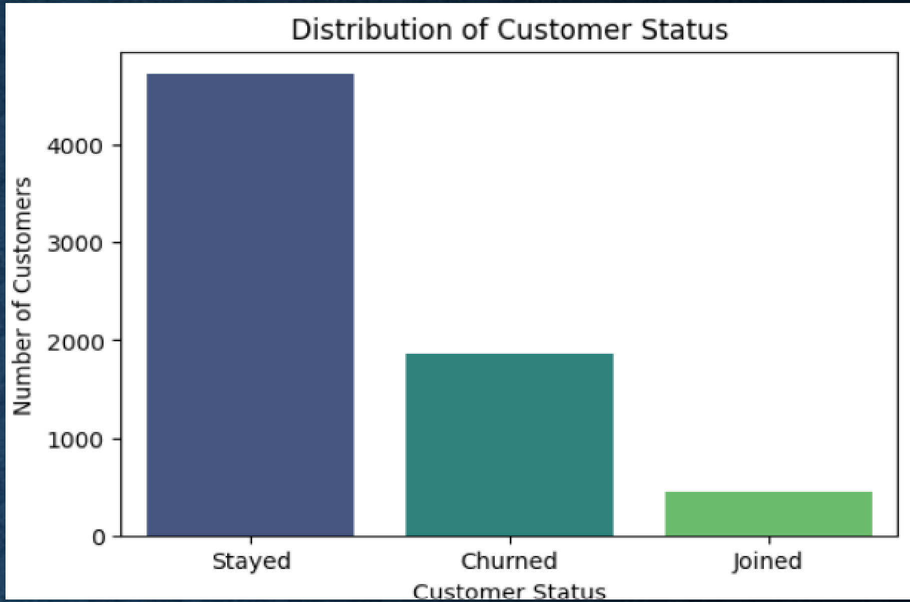
- To explore and analyze telecom customer data using descriptive analytics.
- To develop and evaluate machine learning models for predicting customer churn and churn categories.
- To integrate explainable AI techniques (SHAP and LIME) to enhance model transparency.
- To translate model outputs into meaningful, business-oriented insights.

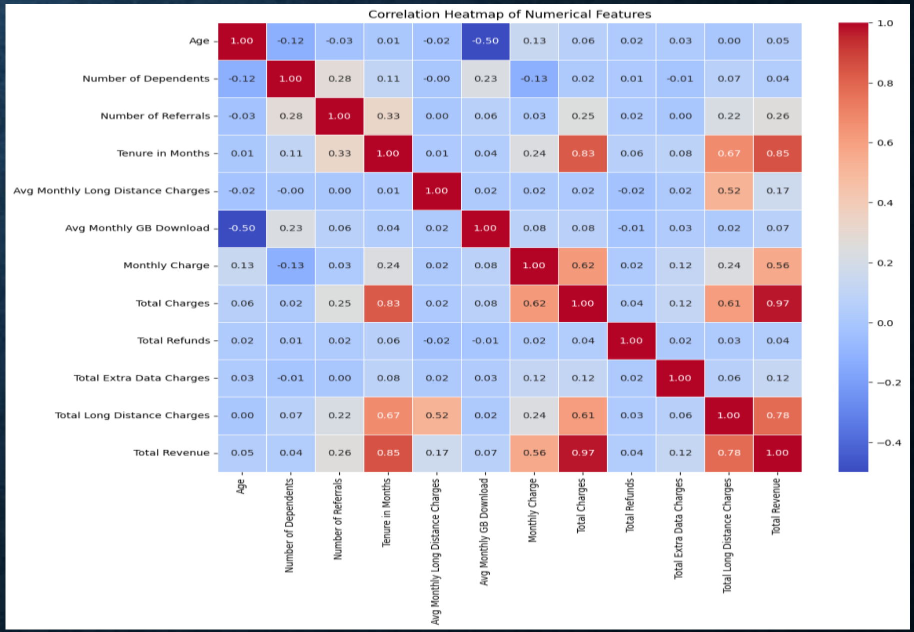
Methodology



Data Information

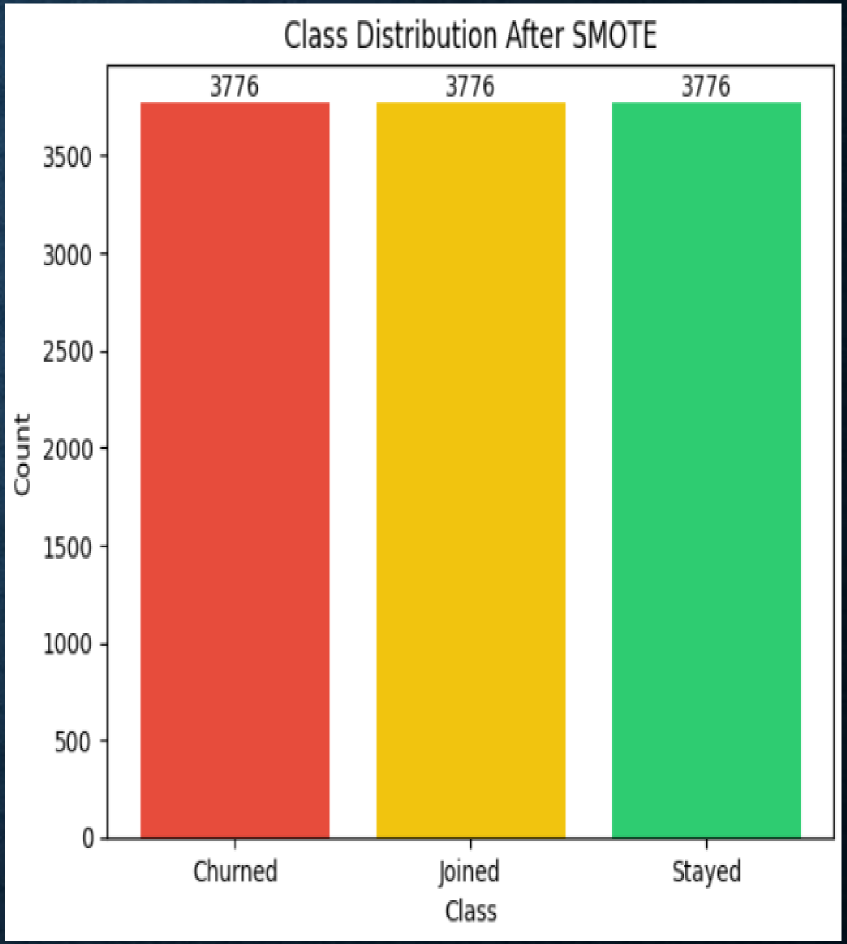
- Dataset: Maven Telecom Customer Churn Dataset (Kaggle – Maven Analytics).
- Size: 7,043 records × 38 features.
- Features: Mix of numerical (tenure, charges) and categorical (contract type, services).
- Targets: Customer Status, Churn Category, and Churn Reason.
- Data Characteristics: Imbalanced dataset; missing values and duplicates handled during preprocessing.





Data Cleaning and Preprocessing

- Initial Dataset: 7,043 rows × 38 columns
- Duplicates: No duplicate records found
- Column Removal: 5 non-informative features removed (Customer ID, City, Zip Code, Latitude, Longitude)
- Missing Values: 30,849 missing values handled using mode (categorical) and median (numerical) imputation
- Encoding: Categorical variables converted using one-hot encoding
- Feature Selection: SelectKBest (k = 20) applied to retain the most relevant features
- Final Dataset for Modelling: 7,043 rows × 20 features
- Class Imbalance: SMOTE applied to balance churn classes

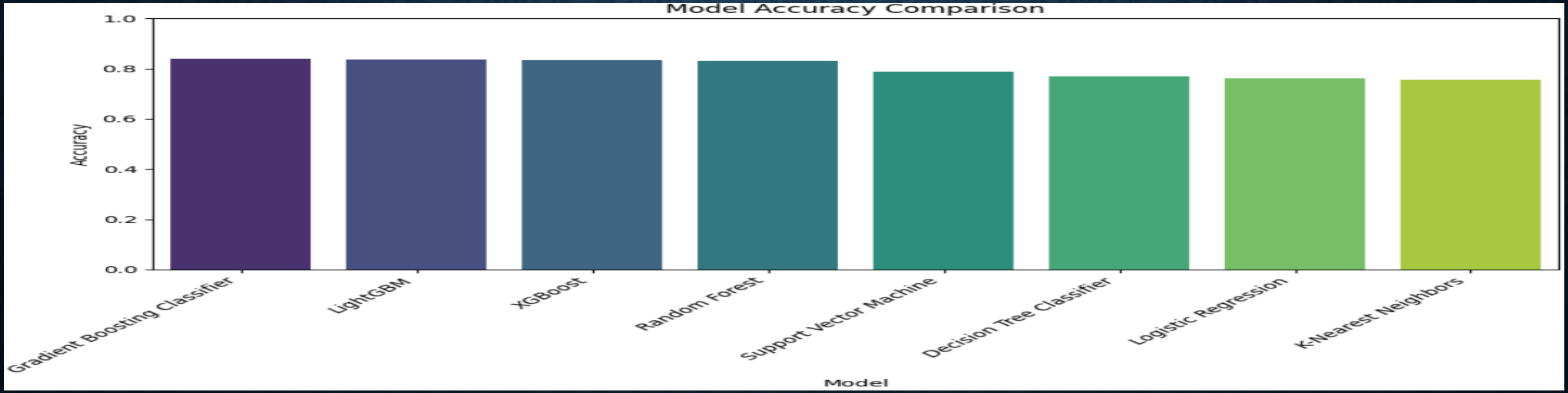


Model selection and Development

- Logistic Regression: Baseline, interpretable model with limited ability to capture non-linear churn patterns.
- Decision Tree: Rule-based model that captured churn logic but showed sensitivity to data variations.
- Random Forest: Ensemble approach that improved stability and handled feature interactions effectively.
- Gradient Boosting: Sequential learning model that captured complex churn relationships.
- XGBoost: Robust model for structured and imbalanced data with strong predictive performance.
- LightGBM: Efficient and fast model with competitive accuracy on the churn dataset.
- SVM: Learned decision boundaries well but was sensitive to scaling and dataset size.
- KNN: Similarity-based model with limited performance on high-dimensional churn data.

Model Accuracy Comparison:

Gradient Boosting Classifier	0.838893
LightGBM	0.836054
XGBoost	0.833215
Random Forest	0.831796
Support Vector Machine	0.787083
Decision Tree Classifier	0.768630
Logistic Regression	0.760823
K-Nearest Neighbors	0.755855



Training And Tuning Models

Training Process:

- Dataset split into **80% training (5,634 records)** and **20% testing (1,409 records)**.
- Stratified sampling applied based on **Customer Status**.

Training Setup:

- Models trained using **20 selected features**.
- **SMOTE** applied on training data to handle class imbalance.

Hyperparameter Tuning:

- **RandomizedSearchCV** used for tree-based and linear models.
- **Cross-validation** applied to ensure robust performance.

Model Saving:

- Best-performing models saved for evaluation and deployment.

Model Evaluation Results

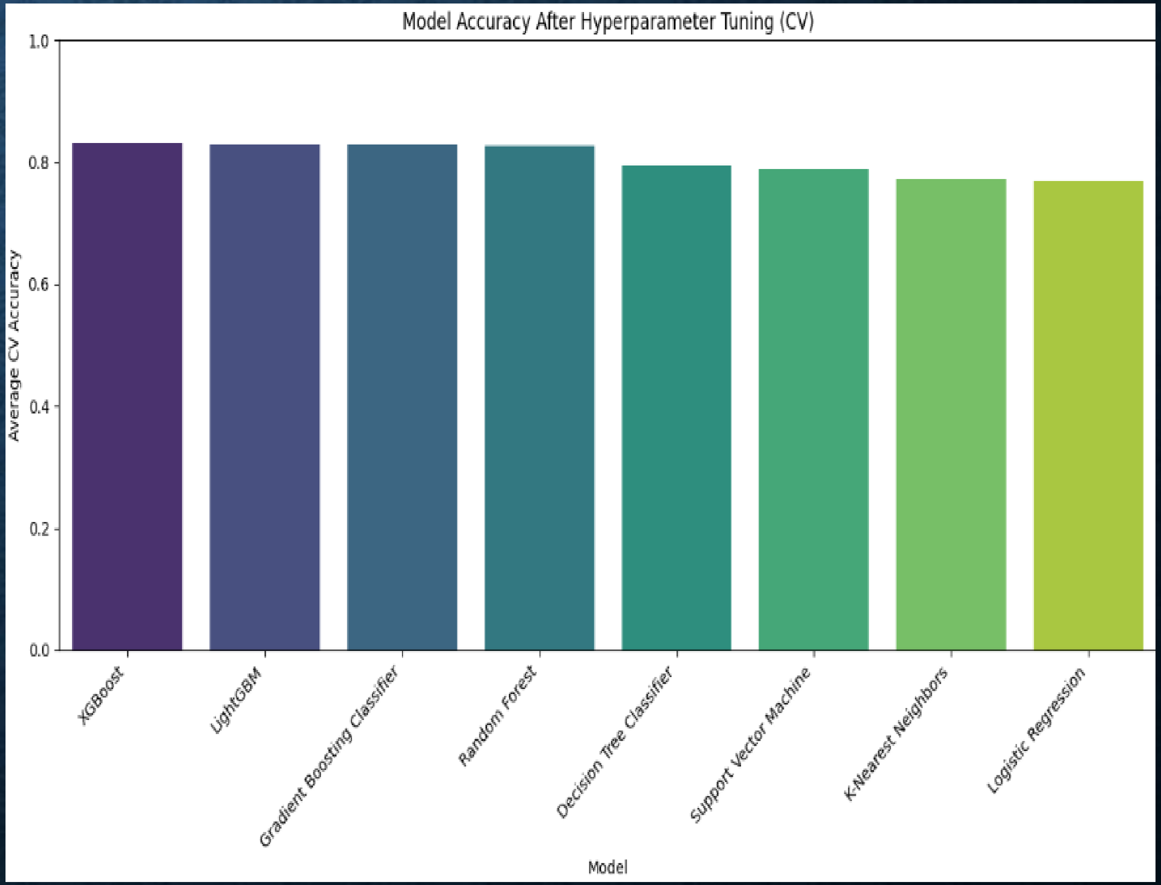
- XGBoost achieved the highest cross-validation accuracy (0.8319) during hyperparameter tuning and was selected as the final model.
- LightGBM (0.8296) and Gradient Boosting (0.8286) showed competitive performance, indicating the effectiveness of ensemble methods on telecom churn data.
- Random Forest (0.8278) also performed well, demonstrating strong stability and robustness.
- Logistic Regression, SVM, and KNN achieved lower accuracies, highlighting limitations of simpler or distance-based models for complex churn patterns.
- Overall, ensemble learning methods consistently outperformed traditional models on the dataset.

Model	Best CV Accuracy	Best Hyperparameters
Logistic Regression	0.7690	solver=liblinear, penalty=l2, C=10.0
Random Forest	0.8278	n_estimators=300, max_depth=10, max_features=log2, min_samples_split=5, min_samples_leaf=1
LightGBM	0.8296	n_estimators=100, learning_rate=0.05, num_leaves=31, max_depth=-1
XGBoost	0.8319	n_estimators=200, learning_rate=0.1, max_depth=6, subsample=1.0, colsample_bytree=0.8
Support Vector Machine	0.7887	kernel=rbf, C=5, gamma=scale
K-Nearest Neighbors	0.7713	n_neighbors=11, weights=distance, p=1
Gradient Boosting Classifier	0.8286	n_estimators=200, learning_rate=0.1, max_depth=3
Decision Tree Classifier	0.7951	max_depth=10, min_samples_split=5, min_samples_leaf=1

Comparison of Tuned Models

- The table and plot compare cross-validated performance of all tuned models.
- Boosting-based models (XGBoost, LightGBM, Gradient Boosting) consistently outperformed other approaches.
- XGBoost achieved the best overall balance of accuracy and stability, leading to its selection as the final model.

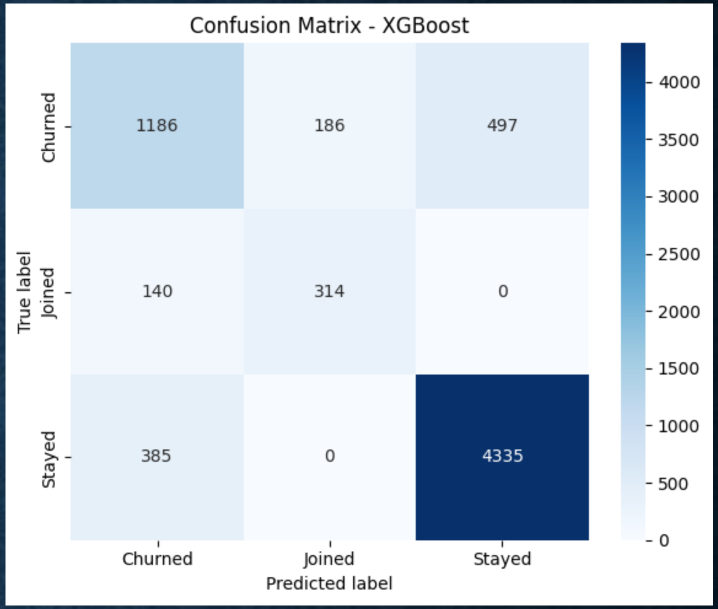
Model	Accuracy	Precision (Churned)	Recall (Churned)	F1-Score (Churned)	ROC-AUC
XGBoost (Final Model)	**0.832**	**0.690**	**0.640**	**0.670**	**0.937**
LightGBM	0.830	0.700	0.630	0.660	0.891
Gradient Boosting	0.829	0.690	0.630	0.660	0.891
Random Forest	0.828	0.680	0.650	0.660	0.884
Decision Tree	0.795	0.640	0.620	0.630	0.833
Support Vector Machine	0.789	0.650	0.670	0.660	0.870
K-Nearest Neighbours	0.771	0.630	0.680	0.650	0.853
Logistic Regression	0.769	0.590	0.760	0.640	0.857

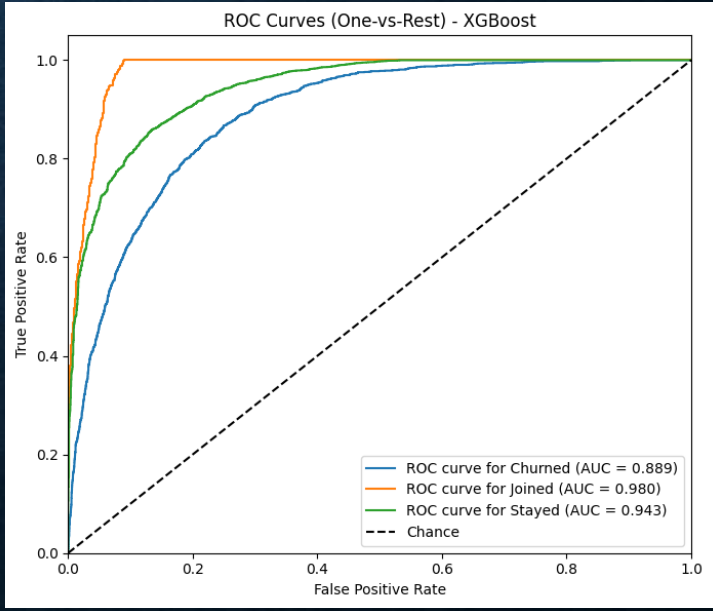


Results and Insights of Best Model

Best Model (XGBoost)

- XGBoost achieved the highest cross-validation accuracy (~83%) after hyperparameter tuning.
- The model showed consistent performance across all customer status classes on the SMOTE-balanced data.
- Most "Stayed" (4,335) and "Churned" (1,186) customers were correctly classified, indicating reliable churn detection.
- ROC–AUC scores showed strong class separation: Stayed: 0.943, Joined: 0.980, Churned: 0.889.
- Overall, XGBoost effectively captured complex churn patterns and was selected as the final model.

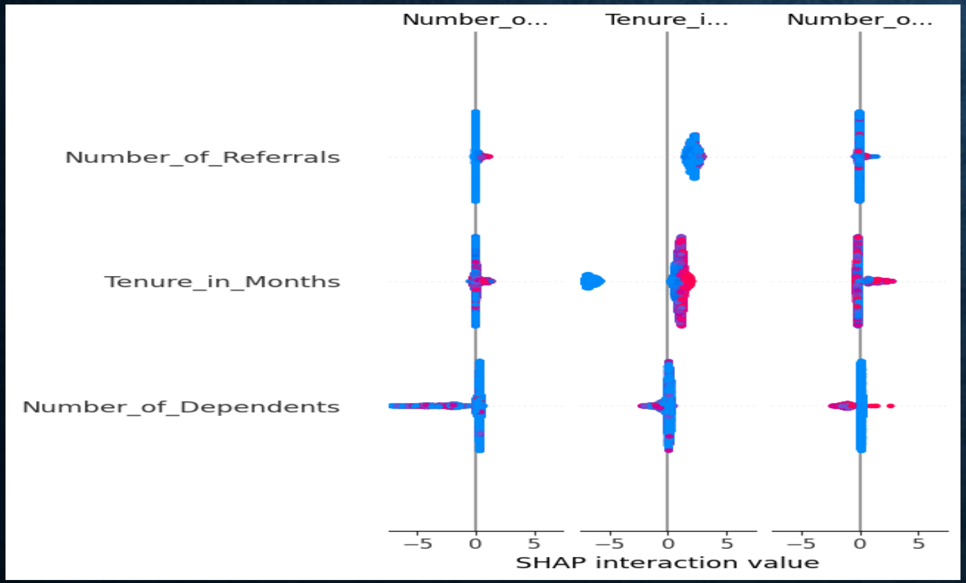




SHAP Global & Local Explanations For Champion Model (XG-BOOST)

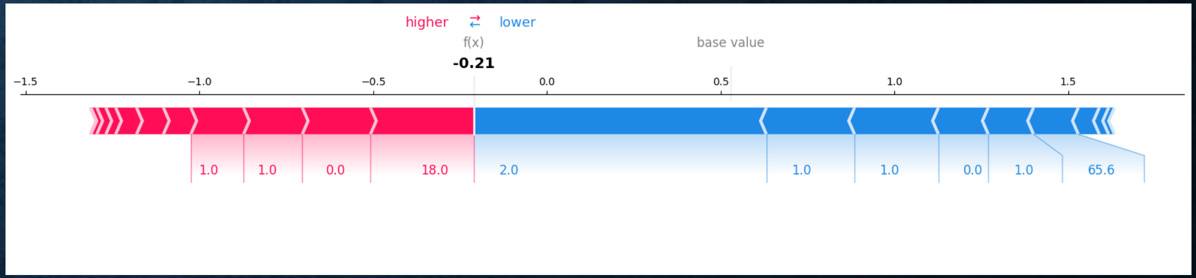
Global Explanation:

- SHAP analysis identifies key features influencing churn predictions across the dataset.
- Tenure in Months, Number of Referrals, and Number of Dependents show strong impact on model decisions.
- Feature interactions indicate how combinations of customer behavior influence churn outcomes.



Local Explanation:

- SHAP force plot explains individual customer predictions.
- Red features push the prediction towards churn, while blue features push it towards non-churn.
- Enables transparent, case-level interpretation of XGBoost predictions.

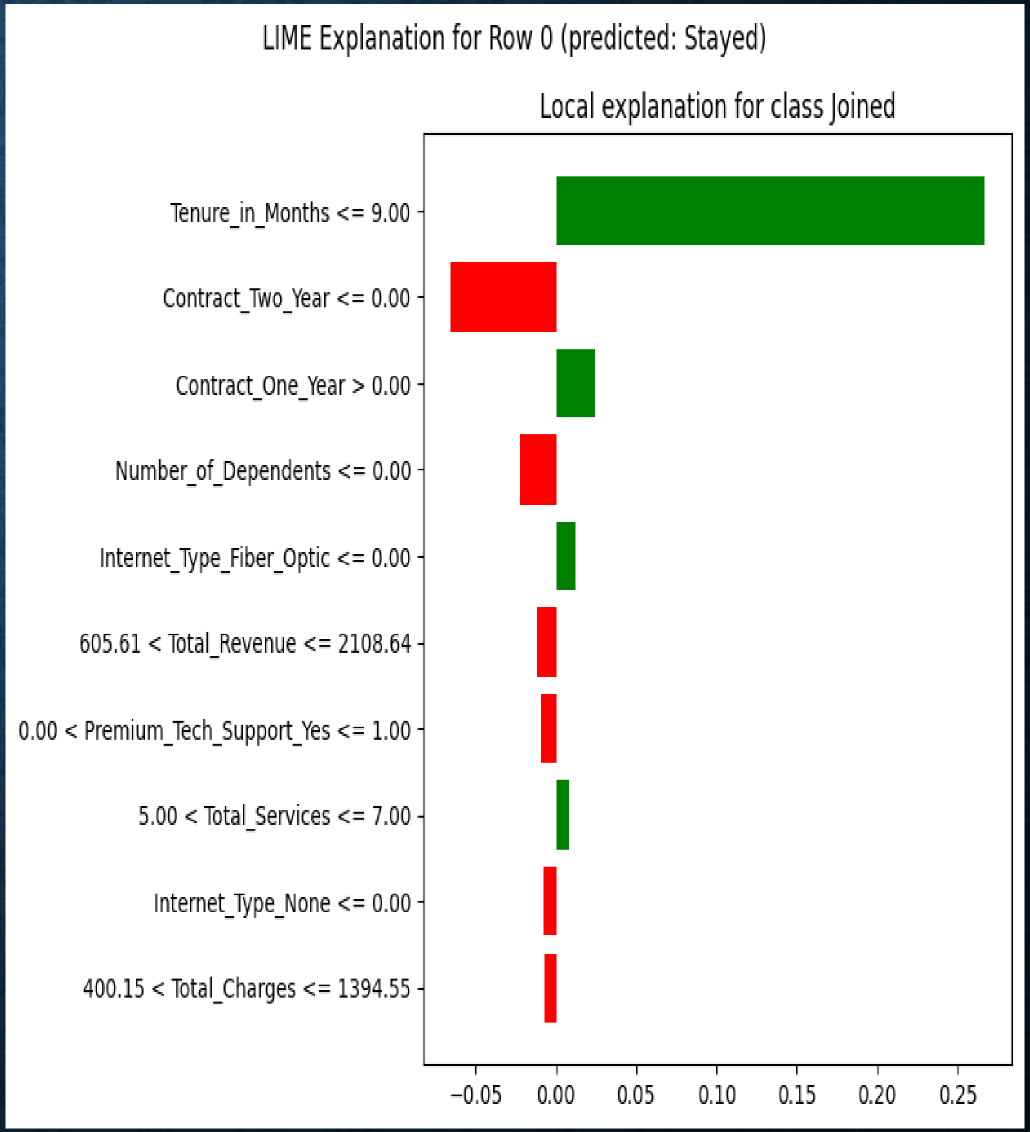


LIME: Local Interpretation of Individual Churn Prediction

- True & Predicted Class: Stayed (≈95% confidence).
- Tenure ≤ 9 months was the strongest factor influencing the prediction.
- One-Year and Two-Year contracts reduced churn risk.
- Dependents and Internet Type had moderate impact.
- Green features support the Stayed prediction; red features increase churn tendency.

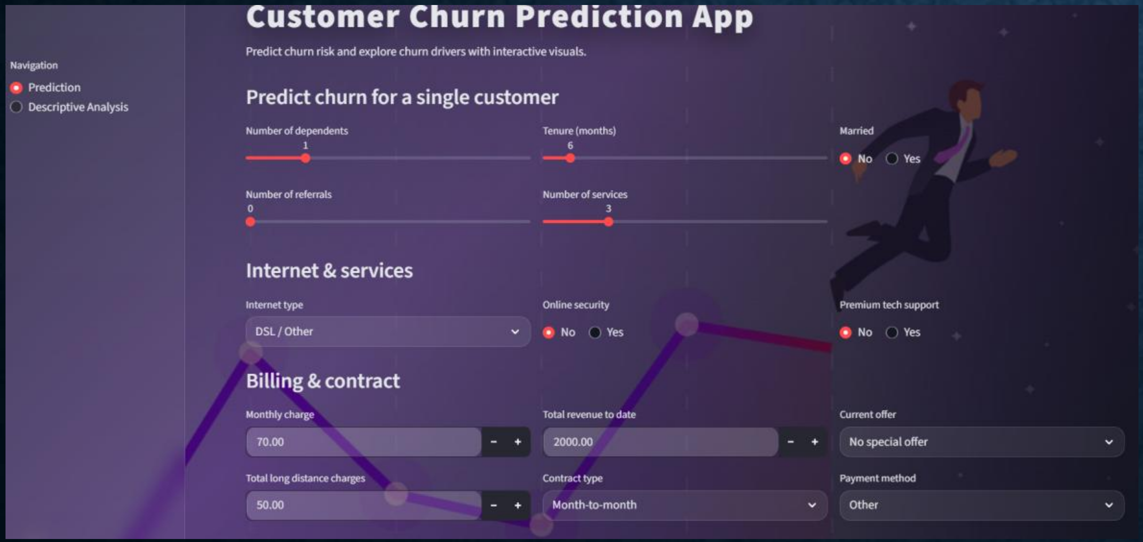
Business Insight:

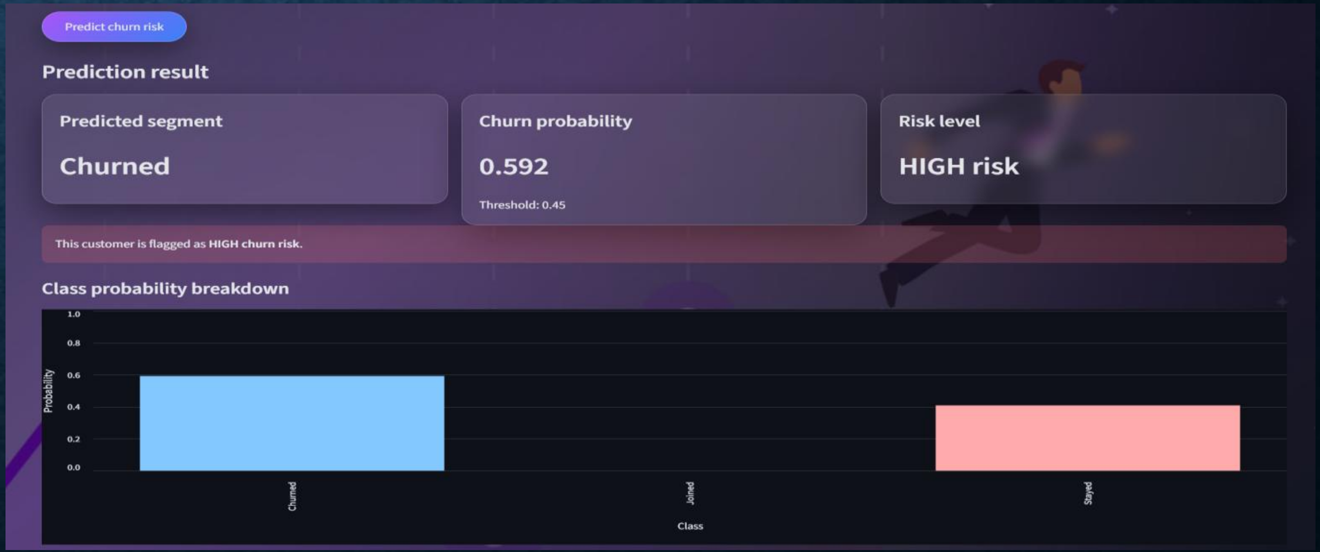
- Customers with stable contracts and longer tenure are less likely to churn.
- LIME supports customer-level, explainable decisions.



Web Application Development & Testing

- Developed an **interactive Streamlit-based churn prediction app**.
- Provides **real-time churn prediction**, probability, and risk level for individual customers.
- Includes **descriptive analytics** to explain *why* customers churn (categories, reasons, segments).
- Visualizes **churn drivers by contract type and customer behavior**.
- Supports **data-driven retention decisions** through explainable insights.





Novelty Contributions

- **End-to-End Pipeline:** Integrated preprocessing, SMOTE balancing, feature selection, and model optimisation for telecom churn prediction.
- **Ensemble-Based Modelling:** Evaluated multiple models and identified **XGBoost** as the most effective approach.
- **Explainable AI Integration:** Combined **SHAP** (global) and **LIME** (local) for transparent, customer-level explanations.
- **Business Interpretability:** Converted model outputs into actionable insights for churn reduction strategies.
- **Real-Time Web Application:** Deployed the trained model as a **Streamlit-based web app** for interactive churn prediction and interpretation.

The key novelty lies in combining robust model optimization with explainable AI to deliver both high performance and business interpretability for telecom churn prediction

Challenges And Solutions

- Challenge: Imbalanced customer status classes*(Stayed 67.0%, Churned 26.5%, Joined 6.5%)*
  Solution: Applied SMOTE to balance classes in the training data.

- **Challenge:** Large number of missing values across service and churn-related features.
  Solution: Used **business-logic based imputation** (e.g., *No Internet*, *No Churn*) and **median imputation** for numerical fields.

- **Challenge:** High-dimensional feature space after categorical encoding.
  Solution: Performed **feature selection (SelectKBest)** to retain the top 20 most informative features.

- **Challenge:** Black-box behavior of high-performing models (XGBoost, LightGBM).
  Solution: Integrated **SHAP (global & local)** and **LIME** for model explainability.

- **Challenge:** Binary churn prediction provides limited business insight.
  Solution: Used **Churn Category** and **Churn Reason** for descriptive analysis to understand *why* customers churn.

- **Challenge:** Default probability threshold not optimal for churn detection.
  Solution: Performed **threshold tuning** for the *Churned* class (optimal threshold = **0.45**).

- **Challenge:** Difficulty in communicating results to non-technical stakeholders.
  Solution: Developed an **interactive Streamlit dashboard** with churn risk scores, probabilities, and visual explanations.

Conclusion And Future Work

Conclusion:

- Built an explainable machine learning framework for telecom churn prediction.
- XGBoost achieved the best performance with strong accuracy and ROC-AUC.
- SHAP and LIME improved transparency and trust in model predictions.
- Descriptive analysis revealed key churn drivers such as contract type, tenure, pricing, and competition.
- The Streamlit dashboard demonstrated real-world applicability.

Future Work:

- Apply time-series models for early churn detection.
- Integrate unstructured data (customer feedback, support logs).
- Use cost-sensitive learning to prioritize high-value customers.
- Explore deep learning models for larger datasets.
- Add fairness and bias analysis for responsible AI deployment.

References

[1] W. Verbeke, K. Dejaeger, D. Martens, J. Hur, and B. Baesens, “New insights into churn prediction in the telecommunication sector,” *European Journal of Operational Research*, vol. 218, no. 1, pp. 211–229, 2012, doi: 10.1016/j.ejor.2011.09.031.

[2] T. Chen and C. Guestrin, “XGBoost: A scalable tree boosting system,” in *Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, San Francisco, CA, USA, 2016, pp. 785–794, doi: 10.1145/2939672.2939785.

[3] S. M. Lundberg and S.-I. Lee, “A unified approach to interpreting model predictions,” in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 30, 2017, pp. 4765–4774.

[4] M. T. Ribeiro, S. Singh, and C. Guestrin, “Why should I trust you? Explaining the predictions of any classifier,” in *Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, San Francisco, CA, USA, 2016, pp. 1135–1144, doi: 10.1145/2939672.2939778.

[5] C. Molnar, *Interpretable Machine Learning*, 2nd ed., 2022. [Online]. Available: https://christophm.github.io/interpretable-ml-book/

THANK YOU