CUSTOMER CHURN PREDICTION USING CLASSIFICATION ALGORITHMS AND SURVIVAL ANALYSIS

August 2024

Introduction to Customer Churn

- **Definition:** Customer churn refers to the loss of clients or subscribers, a critical issue in subscription-based industries.
- **Importance:** Understanding and predicting churn helps businesses to retain customers, saving costs associated with acquiring new ones.
- **Project Scope:** This project focuses on developing predictive models to forecast churn, enabling proactive customer retention strategies.

Objectives and Goals

- **Primary Objective:** Develop a predictive model to identify customers likely to churn, enabling targeted retention efforts.
- **Secondary Goals:** Integrate various data sources (interaction logs, transaction history, demographics) and apply both classification algorithms and survival analysis.
- **Business Impact:** Reduce churn rates, increase customer lifetime value (CLTV), and enhance overall profitability.

Photo by Stephen Dawson on Unsplash

Methodology Overview

- **Data Sources:** Customer interaction logs, transaction history, demographic data.
- **Analytical Techniques:** Classification algorithms (Decision Trees, Logistic Regression) and Survival Analysis (Kaplan-Meier, Cox Proportional Hazards).
- **Data Collection:** Utilized the IBM Watson Telco Customer Churn dataset for building and testing the model.

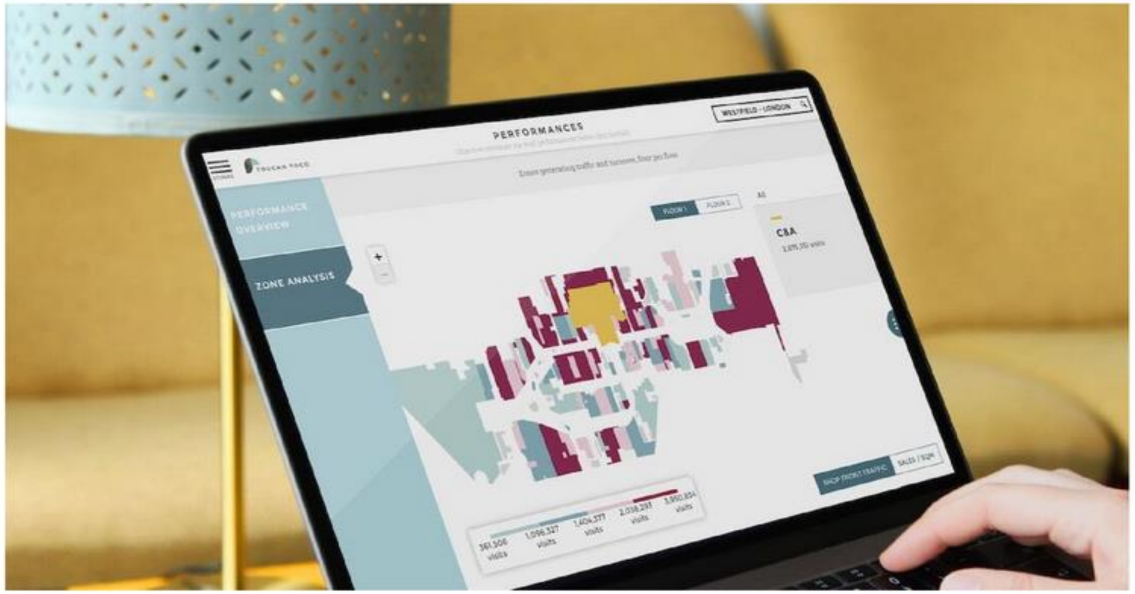

Photo by Adrien WIESENBACH on Unsplash

Model Development & Challenges

- **Models Used:** Developed classification models using Decision Trees, Logistic Regression, and Survival Analysis techniques.
- **Challenges:** Addressed data imbalance with SMOTE and managed feature correlation to avoid multicollinearity.
- **Tuning & Evaluation:** Hyperparameter tuning and cross-validation performed to enhance model performance.

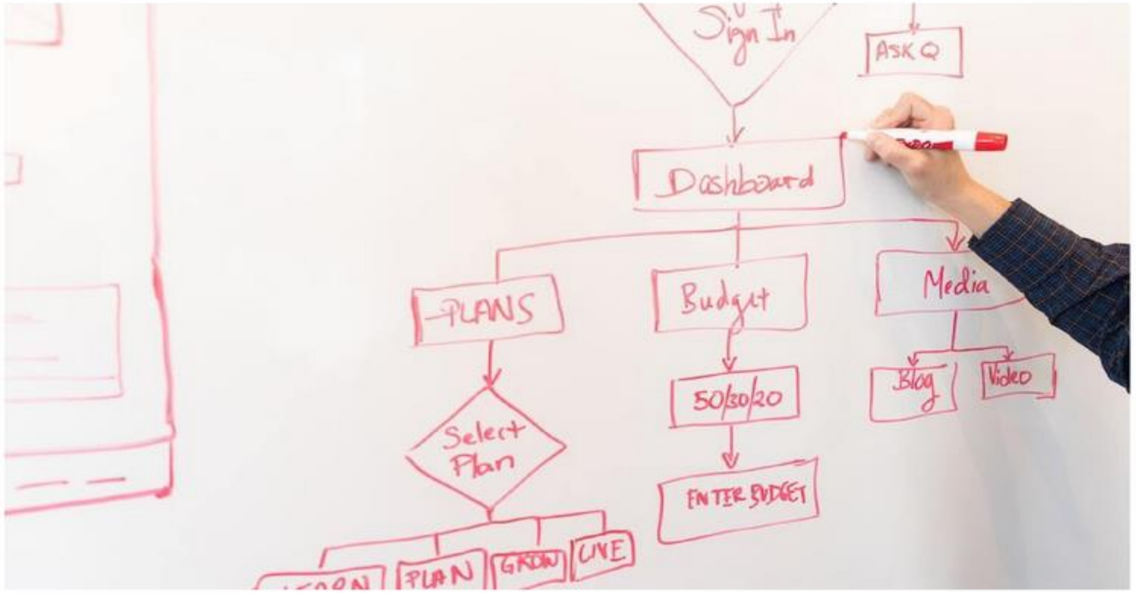

Photo by Christina @ wocintechchat.com on Unsplash

Evaluation Metrics

- **Key Metrics:** Accuracy, Precision, Recall, F1-Score, AUC-ROC for classification; C-Index for survival analysis.
- **Cross-Validation:** K-fold cross-validation (k=10) used to ensure model robustness and prevent overfitting.
- **Business Impact:** Cost-benefit analysis to evaluate model's practical utility in reducing churn and enhancing customer lifetime value.

Photo by Possessed Photography on Unsplash

Implementation & Deployment

- **Model Integration:** The final model will be implemented using Python libraries such as Scikit-learn, Lifelines, and Pandas.

- **Business Application:** A dashboard will be created to visualize churn predictions and key customer metrics for business use.

- **Deployment:** The model will be deployed in a real-time environment for continuous monitoring and prediction of customer churn.

Photo by Slidebean on Unsplash

Expected Outcomes

- **Churn Reduction:** The model is expected to significantly reduce churn rates by accurately identifying at-risk customers.
- **Increased Customer Lifetime Value (CLTV):** Retention strategies informed by the model should increase CLTV, driving long-term profitability.
- **Actionable Insights:** The model will provide interpretable results, guiding targeted marketing and customer retention efforts.

Photo by Carlos Muza on Unsplash

Conclusion & Next Steps

- **Summary:** The predictive model aims to address the challenge of customer churn by leveraging classification algorithms and survival analysis.
- **Future Work:** Further refinement of the model and exploration of additional data sources could enhance predictive accuracy.
- **Business Integration:** Focus on integrating the model into daily business operations to continuously monitor and mitigate churn.

Photo by Mari Helin on Unsplash