Churn Prediction: A Proactive Approach

This project aimed to develop a robust churn prediction model, equipping the organization with valuable insights to enhance customer retention strategies. By leveraging advanced analytics, the model identified key drivers of churn, empowering data-driven decision making.

Andrea Lolli

503035

Bachelor of Science in Artificial Intelligence

Data Preparation: Laying the Foundation

1 Cleaning and Transforming

The dataset underwent a thorough cleaning process, addressing missing values, data type inconsistencies, and other discrepancies to ensure data integrity.

2 Exploratory Data Analysis

In-depth EDA provided valuable insights into the distribution of features, outliers, and relationships within the data, laying the groundwork for feature engineering.

3 Addressing Imbalance

Implemented SMOTE, a powerful technique for balancing the churn classes, enhancing the model's ability to learn from the minority class.

Feature Engineering: Unlocking Insights

Transformations

Features like MultipleLines and internet-related variables were re-encoded to reduce complexity and redundancy, optimizing the model's performance.

New Feature Creation

Developed a novel feature, AutomaticPaymentMethod, which distinguished between automatic and non-automatic payment methods, hypothesized to influence churn.

Streamlined Pipelines

Custom transformers were integrated into a pipeline, ensuring efficient and consistent data preprocessing, a critical step in the model development process.

Model Development: Optimizing for Performance

Feature Selection

Utilized a comprehensive feature selection approach, identifying the most impactful attributes to be included in the model.

Hyperparameter Tuning

Optuna, a powerful hyperparameter optimization framework, was employed to fine-tune the RandomForest model for improved performance.

1

2

3

Model Selection

RandomForest was chosen for its ability to handle both categorical and continuous data, as well as its robustness to overfitting.

Model Evaluation: Ensuring Robustness

Comprehensive Metrics

The model was evaluated using a suite of performance metrics, including accuracy, precision, recall, F1-score, and ROC AUC.

Cross-Validation

Rigorous cross-validation was conducted to ensure the model's stability and reliability, confirming consistent performance across different data subsets.

Precision-Recall Analysis

The precision-recall curve highlighted the model's strong ability to predict the minority churn class, a critical consideration for this business problem.

Feature Importance

Analysis of feature importance revealed key drivers of churn, such as contract duration, tenure, and monthly charges, guiding targeted retention strategies.

Insights and Business Impact

Contract Type

Contract duration and type emerged as significant predictors of churn, enabling tailored interventions for at-risk customers.

Tenure

Tenure was identified as a critical factor, suggesting the need for personalized retention strategies for customers with shorter tenures.

Payment Method

The model revealed that payment method, particularly automatic versus non-automatic, influences churn, guiding targeted customer outreach.

Strategic Recommendations

1

Targeted Interventions

Leverage the model's insights to implement tailored retention strategies, such as personalized offers or loyalty programs, for customers identified as high-risk.

2

Customer Segmentation

Segment customers based on the model's key predictors, enabling focused attention and resource allocation towards the most vulnerable segments.

3

Proactive Monitoring

Continuously monitor the model's performance and update it with new data to ensure the organization stays ahead of evolving churn patterns.

Conclusion: A Comprehensive Approach

This churn prediction project demonstrates a robust framework for identifying customers likely to churn, leveraging advanced analytics and a systematic approach. The insights generated can significantly aid strategic decision-making, optimizing customer retention efforts and driving business growth.

Limitations and Future Work

Limitations

The model's performance is dependent on the quality and completeness of the dataset. Incorporating additional data sources could further enhance the predictive capability.

Explainability

While RandomForest provides interpretable feature importance, exploring more explainable models could offer deeper insights into the underlying drivers of churn.

Future Enhancements

Integrating the model into a real-time monitoring system and continuously updating it with new data could enable proactive and adaptive churn management.