**Customer Churn Prediction: Reducing Revenue Loss Through Machine Learning**

**Introduction**

Telecommunications companies face significant revenue loss from customer churn. For this project, I developed a predictive model to identify at-risk customers 60 days before likely cancellation, enabling proactive retention strategies. This initiative combined data science techniques with business strategy to address a critical revenue challenge.

**Problem Statement**

The telecommunications company was experiencing a 15% annual customer churn rate, resulting in approximately $3.2M in lost revenue. Traditional retention methods were purely reactive—only engaging customers after they requested cancellation, with just a 12% success rate for winback campaigns. The company needed to shift to a proactive approach by identifying at-risk customers before they decided to leave.

**Methodology**

**1. Data Collection & Integration**

* Sourced customer demographic data, service usage patterns, billing information, and support interactions
* Integrated data from CRM, billing systems, and product telemetry into a unified dataset
* Created a comprehensive view of 20,000+ customers over an 18-month period

**2. Data Preparation & Exploration**

* Cleaned data and handled missing values using imputation techniques
* Performed exploratory analysis to identify patterns in churned vs retained customers
* Analyzed correlation between service usage metrics and churn probability
* Discovered key indicators including service interruptions, billing disputes, and usage decline patterns

**3. Feature Engineering**

* Created features capturing usage trends, service interruptions, and support ticket history
* Developed customer segmentation based on behavior patterns
* Generated time-based features to capture changes in usage over time
* Engineered features around payment behavior (late payments, billing disputes)

**4. Model Development**

* Implemented multiple classification algorithms (Logistic Regression, Random Forest, XGBoost)
* Performed hyperparameter tuning to optimize model performance
* Evaluated models using precision, recall, and F1 score with emphasis on correctly identifying churners
* Selected final XGBoost model based on superior performance metrics and interpretability

**5. Business Implementation**

* Identified key churn indicators and their relative importance
* Developed scoring system for customer churn risk categorization
* Created framework for continuous model retraining and validation
* Designed intervention strategies based on risk scores and customer segments

**Technical Tools Used**

* **Python**: Primary programming language
* **Scikit-learn**: Core machine learning implementation
* **XGBoost**: Advanced gradient boosting for the final model
* **Pandas/NumPy**: Data manipulation and analysis
* **SQL**: Data extraction and integration
* **Tableau**: Visualization and business dashboard creation

**Challenges Overcome**

* **Data Silos**: Overcame organizational barriers to integrate data from multiple systems
* **Class Imbalance**: Implemented SMOTE and class weighting techniques to address the imbalanced dataset (85% retained vs. 15% churned)
* **Feature Selection**: Reduced initial 200+ features to 45 most predictive variables without losing model performance
* **Model Interpretability**: Created simplified explanations and visualizations to help business stakeholders understand complex model outputs

**Results & Business Impact**

* Achieved 83% accuracy in predicting customer churn 60 days in advance
* Identified that service outages and declining usage were strongest predictors
* Enabled targeting of high-risk customers with personalized retention offers
* Implemented A/B testing framework to measure intervention effectiveness
* Projected 12% reduction in churn rate through proactive intervention
* Estimated annual savings of $1.8M based on initial pilot program results

**Key Learnings & Skills Demonstrated**

* End-to-end machine learning project management
* Cross-functional collaboration with marketing, sales, and customer service teams
* Feature engineering for time-series behavioral data
* Model validation and monitoring in production environments
* Translating technical insights into actionable business recommendations

Would you like me to expand on any specific section or add any additional elements to make this portfolio piece even stronger?