Machine_Learning_phase_3_project Public

1 Branch   0 Tags

Go to file   Go to file   Add file +   About   Code   ⋯

Sam-Kossie Add README   4ed3f3f · 5 minutes ago   🕐 5 Commits

📄 Phase 3 Project Pre...	Presentation addition	9 minutes ago
📄 README.md	Add README	5 minutes ago
📄 bigml_59c2883133...	Project final commit	15 minutes ago
📄 notebook1.ipynb.ip...	Project final commit	15 minutes ago

Machine_Learning_phase_3_project

📖 Readme
〰️ Activity
☆ 0 stars
👁 0 watching
ᕦ 0 forks

Releases

No releases published
Create a new release

Packages

No packages published
Publish your first package

Languages

📖 README   ✏️ ☰

Machine_Learning_phase_3_project

Project Title :SyriaTel Customer Churn Prediction

Business Understanding

Stakeholder: Customer Retention Department of SyriaTel.

Business Problem: SyriaTel is experiencing customer loss (churn) to competitors. High churn rates directly impact revenue and market share. Because retaining an existing customer is more cost-effective than acquiring a new one, SyriaTel needs a predictive model to identify customers likely to leave.

Data Understanding

The dataset consists of 3,333 records with 21 features.

Target: churn (Boolean: True/False).

Key Features: Account length, usage minutes (day, evening, night, international), service plans, and customer service calls.

Modelling

Two classification models were developed and compared:

Logistic Regression

- Simple interpretable baseline model
- Used to identify key features driving churn
- Coefficients revealed the most important predictors

Decision Tree Classifier

- Max depth limited to 5 to prevent overfitting
- Better at capturing non-linear relationships
- Achieved 92.2% accuracy on test data

Both models were trained on 70% of the data and validated on the remaining 30%. Feature engineering and preprocessing included handling categorical variables and scaling numerical features where appropriate.

Evaluation

Decision Tree Performance:

- **Accuracy**: 92.2%
- **Precision (Churn class)**: 78%
- **Recall (Churn class)**: 67%
- **F1-Score (Churn class)**: 72%

Key Findings:

- Customer service calls emerged as a top predictor of churn
- High daytime usage minutes showed strong correlation with churn
- International plan subscribers with low usage showed higher churn rates
- ROC-AUC curves demonstrated good model discrimination ability

The model performs well at identifying non-churners but has room for improvement in recall for the churn class, meaning some at-risk customers may still be missed.

Limitations

1. **Class Imbalance**: The dataset has fewer churned customers, which may affect the model's ability to learn churn patterns effectively

2. **Recall Trade-off**: Current recall of 67% means approximately one-third of churning customers might not be identified

3. **Geographic Data Excluded**: State information was not fully utilized, potentially missing regional patterns

4. **Temporal Factors**: The dataset represents a snapshot in time and doesn't capture seasonal trends or customer lifecycle stages

5. **Feature Limitations**: Some potentially important factors like competitor pricing, customer demographics, or service quality metrics are not available in the dataset

Recommendations

Based on model insights and evaluation results:

1. Service Call Intervention Program

Implement a "3-call threshold" alert system where customers who contact support three or more times automatically receive follow-up from a senior specialist. This addresses the strongest churn predictor identified by the model.

2. Daytime Usage Loyalty Program

Create incentive programs for high daytime users, such as:

- Volume-based discounts
- Free additional daytime minutes for loyal customers
- Flexible pricing plans for heavy daytime users

3. International Plan Optimization

Proactively review international plan subscribers with low usage and:

- Suggest plan downgrades to better match their needs
- Provide education about alternative calling options
- Offer trial periods for plan changes

4. Enhance Model Performance

- Collect additional data to improve recall through better class balance
- Implement cost-sensitive learning to prioritize churn detection
- Consider ensemble methods (Random Forest, XGBoost) for improved predictions

Conclusion

This project successfully developed a customer churn prediction model for SyriaTel with 92.2% accuracy. The model identified customer service calls, daytime usage patterns, and international plan issues as key drivers of churn. While the model provides valuable insights for targeted retention efforts, there is opportunity to improve recall through advanced modeling techniques and additional data collection. Implementing the recommended interventions could significantly reduce churn rates and improve customer satisfaction, ultimately protecting SyriaTel's revenue and market position.