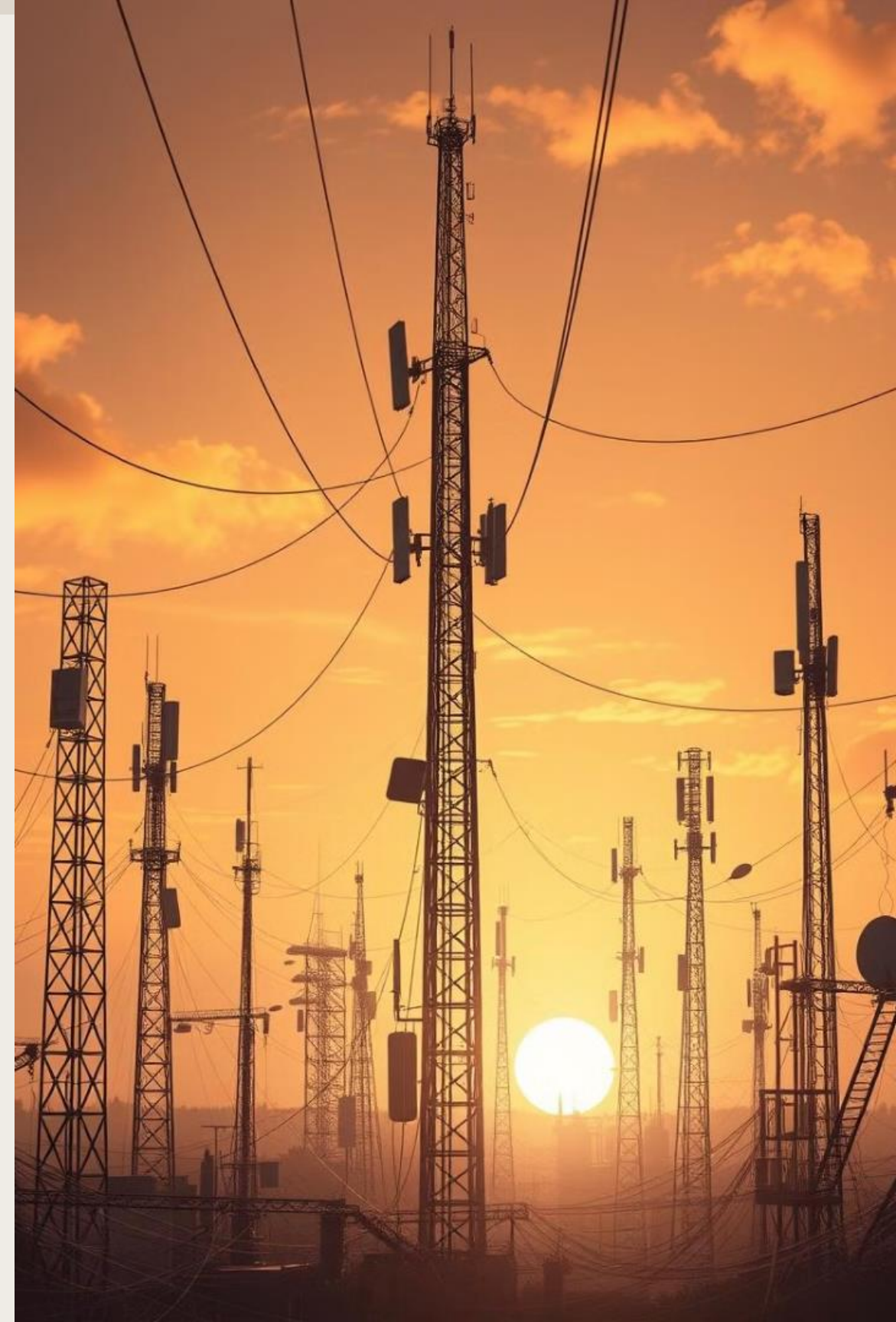


Comparative Analysis of Logistic Regression and Decision Tree Models for Predicting Customer Churn

SyriaTel faces the challenge of customer churn, losing revenue, and impacting profitability.

 by Keziah Gicheha.





Data Understanding

The data for this project was sourced from Kaggle and represents customer information for a telecommunications company.

1 Customer Demographics

Area Code, International Plan, Voice Mail Plan

2 Usage Statistics

Total Day Minutes, Total Day Calls, Total Day Charge

3 Customer Service Interaction

Customer Service Calls

4 Target Variable

Churn: Indicates whether the customer churned (True) or stayed (False).



Data Analysis Questions

The data analysis questions aim to identify factors that indicate customer churn and predict future churn behavior.

Factors

Identify features or behaviors that strongly correlate with customers leaving the service.

Prediction

Build a predictive model to proactively identify at-risk customers.

Accuracy

Understand the model's accuracy to gauge the reliability of predictions.

Customer Segmentation

Segment customers based on their churn risk for personalized retention strategies.



Modeling Approach

The modeling approach involves data preprocessing, model selection, training, and evaluation.

1

Data Preprocessing

Handle missing values, outliers, and incorrect data entries.

2

Feature Engineering

Create new features that might help the model.

3

Scaling

Standardize or normalize numerical features.

4

Balancing the Dataset

Apply techniques like SMOTE to balance the class distribution.

Model Selection

Two models were selected: Logistic Regression and Decision Tree.

Logistic Regression

- Aimed for simplicity and interpretability
- A method to predict outcome-based input features

Decision Tree

- Chosen for its interpretability and ability to handle non-linear relationships.
- A model that makes decisions by splitting data into branches

Goal: Identify which customers are at risk of churning.

Model Evaluation

1. Accuracy

•Decision Tree:

- Training Accuracy: **1.0000** (Overfitting)
- Test Accuracy: **0.91**
- *Summary:* Good generalization despite overfitting.

•Logistic Regression:

- Test Accuracy: **0.53**

•**Decision Tree** outperforms **Logistic Regression** in generalizing to new data, making it more reliable for real-world predictions.

. Precision

•Decision Tree:

- Positive Class Precision: **0.6456**
- *Summary:* Moderate accuracy in identifying true positives.

•Logistic Regression:

- Positive Class Precision: **0.19**
- *Summary:* High rate of false positives, low precision.



Model Evaluation Continued

3. Recall

•Decision Tree:

- Positive Class Recall: **0.7133**
- *Summary*: Fairly good at detecting positives.

•Logistic Regression:

- Positive Class Recall: **0.73**
- *Summary*: Effective at identifying positives, but sacrifices precision

F1-Score

•Decision Tree:

- F1-Score (Positive Class): **0.6777**
- *Summary*: Balanced trade-off between precision and recall.

•Logistic Regression:

- F1-Score (Positive Class): **0.30**

Summary: Weaker performance due to low precision.

- Decision Tree: Better overall performance; needs tuning to reduce overfitting.
- Logistic Regression: Underperforms due to class imbalance; consider data resampling or different model.





Predictive Recommendations

Use Decision Tree for Immediate Predictions

Implement the decision tree model for real-time churn predictions, as it offers higher accuracy and recall for identifying customers at risk of churning.



Monitor and Adjust Model Performance:

Continuously monitor the decision tree's predictions and update the model periodically to account for new data trends, reducing the risk of overfitting.



Combine Models for Improved Predictions

Use an ensemble approach by combining the decision tree and logistic regression models to leverage the strengths of both

Conclusion

The decision tree model offers a clear advantage in interpretability, showing how different features influence churn decisions.

Recommendations

1

Leverage the Decision Tree Model: Given its strong performance on the test set

2

Target High-Risk Customers: Focus retention efforts on the customers

Refine Marketing Strategies: Use the insights from the model to tailor marketing and retention strategies,

3

Retention Strategies

Use model insights to identify at-risk customers.

Ongoing Evaluation

Regularly update and monitor the model.

Customer thurn rediction

Customer churn, corfor the predicction on mail, election is l
anpiorder pedicating inreasdings; ous most for gent pomders.
ss.
fines volemcement and allions ide: this gansdctepient conter
accommaned uustomer remontatiof fell and salec thair lvrnux,
amest other, reperiation.
This are bate tinger recommendtions.
escers: Under curensersatioer Churn and leers fory the
tincal excriative node regested and recomptiorts.

Next Steps

The decision tree model offers a clear advantage in interpretability, showing how different features influence churn decisions.

Customer thurn rediction



1

Recommendations

Model Optimization: Apply techniques like pruning or regularization to the decision tree to reduce overfitting and improve generalization to new data.

2

Data Resampling

under sampling to address class imbalance and improve the performance of the logistic regression model.

3

Cross-Validation

Conduct cross-validation to assess model stability and generalizability, ensuring that the decision tree model performs consistently across different subsets of the data.

Thank you

