

Churn Prediction Model Evaluation

Conclusion

In this analysis, I applied machine learning techniques to predict high-value customers at risk of churning. The analysis involved the use of:

Random Forest: This model was selected for its effectiveness in predicting churned customers based on sensitivity.

PCA : This was used for dimensionality reduction to simplify the feature space and improve model performance.

NearMiss : This technique was used to address the imbalance in the dataset.

Logistic Regression: This method helped identify the key features contributing most to customer churn.

Focus

The evaluation of the models was based primarily on Sensitivity and Accuracy.

Model Selected: Random Forest

Random Forest was selected as the best model due to its high Sensitivity (0.814249) in predicting churned customers.

Top Features :

arpu_7,
std_og_t2m_mou_8,
std_og_t2t_mou_8,
std_ic_mou_8,
vol_2g_mb_8,
spl_ic_mou_8,
vol_3g_mb_8,
total_rech_num_8,
sep_vbc_3g,
last_day_rch_amt_8,
loc_ic_t2m_mou_8

Business Recommendations

Based on the analysis, it is recommended that the company should focus on the following action phase features to reduce churn:

Incoming and Outgoing Call Minutes of Usage

Data Usage (2G and 3G)

Final Thoughts

The Random Forest model has shown promising results, especially in terms of sensitivity. By focusing on the most relevant features and understanding the patterns associated with churn, the company can take proactive measures to retain high-value customers, ultimately contributing to long-term revenue growth.