Customer Churn Prediction

A Data Analytics Project with Cognos

Done by:

Senthil Kumar J - 2021115098
Shruthi B - 2021115106
Yalini Kumar - 2021115122
Yuthikshaa - 2021115123
Kattu Bava K S - 2021115308

DESIGN THINKING:

Designing a solution to improve customer churn prediction accuracy involves several key steps, including data preparation, feature engineering, advanced machine learning techniques, model selection, and evaluation. Here's a high-level design for such a solution:

1. **Data Collection and Preprocessing:**
    - Gather historical customer data, ensuring it's comprehensive and wellstructured.
    - Clean the data by addressing missing values and outliers. Normalize or scale numerical features.
    - Encode categorical variables, if present, using techniques like one-hot encoding or label encoding.
2. **Feature Engineering:**
    - Create new features that capture meaningful customer behaviours and interactions, such as:
        - Time-based features (e.g., days since last purchase, days as a customer)
        - Aggregated statistics (e.g., average transaction amount, frequency of interactions)
        - Customer segmentation (e.g., grouping customers by behaviour or demographics)
3. **Data Splitting:**
    - Split the dataset into training, validation, and test sets, typically in a 70-15- 15 or similar ratio.
4. **Advanced Machine Learning Techniques:**
    - Explore advanced machine learning techniques to improve prediction accuracy, including ensemble methods:
        - Random Forest: Ensemble of decision trees known for robustness and accuracy.

    - Gradient Boosting (e.g., XGBoost, LightGBM, CatBoost): Boosted tree algorithms for high performance.
    - Stacking: Combining multiple models to create a meta-model for better accuracy

5. **Model Training and Hyper Parameter Tuning:**
    - Train and validate multiple models using the training and validation sets.
    - Utilize techniques like cross-validation for robust model selection and hyperparameter tuning.
    - Optimize models for high precision, recall, or an F1-score, depending on business objectives.
6. **Model Evaluation and Interpretability:**
    - Evaluate model performance on the test dataset using appropriate metrics (e.g., accuracy, precision, recall).
    - Consider using techniques like SHAP values or feature importance analysis for model interpretability.
7. **Deployment and Real Time Prediction:**
    - Deploy the best-performing model into a production environment.
    - Set up a mechanism for real-time prediction based on new customer data
8. **Monitoring and Maintanence:**
    - Continuously monitor model performance in the production environment.
    - Implement retraining schedules to adapt to changing customer behavior and market dynamics.
9. **Actionable Insights and Feedback Loop:**

- Use churn predictions to take proactive actions for customer retention, such as targeted marketing campaigns or personalized offers.
- Collect and incorporate feedback on the effectiveness of these strategies to refine the approach.

10. **Documentation and Reporting:**
    - Document the entire process, including data preprocessing, model details, evaluation results, and deployment procedures.
    - Provide regular reports to stakeholders on churn prediction performance and its impact on customer retention and revenue.