**Ideation Phase**

**Defining the Problem Statements**

|  |  |
| --- | --- |
| **Date** | **27-09-2023** |
| **Team ID** | **3923** |
| **Project Name** | **Customer churn prediction in E-commerce by Gradient Boost.** |

**Customer churn prediction by Gradient Boost**

**Problem Definition and Design Thinking**

**Introduction**

In this work, data mining is used to forecast client attrition in an online meal ordering business based in Tehran. Gradient Boosted Trees (GBT), which achieve an accuracy of 86.90%, are shown to be the most efficient method when comparing other approaches. The importance of customer churn prediction for business growth is emphasised in the study, as is the function of data mining in comprehending and maintaining clients.

In this document, we will outline the problem statement, the steps involved in solving it, and the design thinking approach that will guide our project.

**Problem Statement**

Objective: Develop a model that can predict customer churn in e-commerce by Gradient Boost with a high level of accuracy.

Data: Data collected from an online food ordering service in Tehran, Iran. The dataset includes attributes such as city, area, orders count, vendor type, payment type, order source, distance, delivery time, and others. Gradient Boosted Trees (GBT) is chosen as the primary classification method. Other methods include k-NN, Naïve Bayes, Decision Tree, Random Forest, and Rule Induction.

**Key Challenges:**

1. Data Quality: Ensuring the dataset is clean, complete, and free of errors.

2. Feature Selection: Identifying the most relevant features for accurate churn predictions.

3. Model Selection: Choosing the appropriate classification method(s) for the task.

4. Model Evaluation: Evaluating the model's performance using appropriate metrics.

5. Deployment: Creating a user-friendly interface or API for end-users to make predictions.

**Design Thinking Approach**

**Empathize:**

This study empathizes with the critical problem of customer churn in the online food service industry in Tehran, Iran. It recognizes the business imperative to predict and address customer churn, highlighting its impact on profitability. Utilizing data mining techniques, the research aims to provide insights into customer behavior, offering businesses a strategic tool for customer retention.

**Actions:**

- Conduct surveys or interviews with potential users to gather their perspectives.

- Focus on Customer Relationship Management (CRM) strategies for customer retention.

- Seek feedback from domain experts in e-commerce .

**Define:**

Based on our understanding of the problem and the users' needs, we will define clear objectives and success criteria for our project.

**Objectives:**

- Employ various classification methods, including Gradient Boosted Trees, to predict customer churn.

- Create a user-friendly web application for users to input customer details and receive churn predictions.

**Ideate:**

Brainstorm potential solutions and approaches to address the problem. This phase involves thinking creatively and considering various algorithms and techniques for customer churn prediction.

**Actions:**

- Explore different classification models such as k-NN, Naïve Bayes, Decision Tree, Random Forest, and Rule Induction.

- Experiment with feature engineering techniques to enhance model performance.

- Consider incorporating external data sources (e.g., market trends, location-based data) to improve predictions.

**Prototype**

Create a prototype of the model and the user interface for customer churn prediction.

**Actions:**

- Develop a Jupyter Notebook or Python script for data pre-processing, model training, and evaluation.

- Create a simple web interface using tools like Flask or Django to allow users to input customer details.

- Test the prototype with a subset of the dataset to ensure it meets performance objectives.

**Test**

Evaluate the model's performance using appropriate metrics and gather feedback from users.

**Actions:**

- Split the dataset into training and testing sets.

- Train the model on the training set and evaluate it on the testing set.

- Use metrics such as F1 Score, Matthews Correlation Coefficient (MCC), Log Loss, and Precision at K (P@K) to assess model performance.

- Collect user feedback on the web interface for usability and accuracy.

**Implement**

Once the prototype meets the defined objectives and receives positive feedback, proceed with full implementation.

**Actions:**

- Train the final model on the entire dataset.

- Deploy the model as part of a production-ready web application.

- Conduct thorough testing to ensure the application is robust and user-friendly.

**Iterate**

Continuous improvement is essential. Gather user feedback and iterate on the model and interface to enhance accuracy and usability.

**Actions:**

- Monitor the model's performance and retrain it periodically with updated data.

- Address user feedback and make necessary improvements to the web interface.

- Stay informed about advancements in e-commerce for potential enhancements.

**Conclusion**

In this document, we've outlined our approach to solving the problem of customer churn prediction usingGradient Boost. We've defined the problem, identified key challenges, and laid out a design thinking approach that involves empathizing with users, defining objectives, ideating potential solutions, prototyping, testing, implementing, and iterating.

Our ultimate goal is to develop an accurate and user-friendly solution that provides valuable insights for both customers and service providers in e-commerce. By following this structured approach, we aim to create a reliable tool that contributes positively to the e-commerce industry.