**Ideation Phase**

**Defining the Problem Statements**

|  |  |
| --- | --- |
| **Date** | **27-09-2023** |
| **Team ID** | **3923** |
| **Project Name** | **Telco Customer Churn Prediction using Decision Trees and Logistic Regression.** |

**Telco Customer churn prediction using Decision Trees and Logistic Regression**

**Problem Definition and Design Thinking**

**Introduction**

Utilizing data mining and machine learning techniques, specifically logistic regression and decision trees, the study proposes a model for identifying high-risk churn customers. The system, implemented in R programming, offers options for performance analysis, testing, and training on diverse datasets. The proposed approach aims to provide actionable insights for telecom companies to strategically retain customers, ultimately reducing revenue losses associated with churn.

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 Telecom by Decision trees and logistic regression with a high level of accuracy.

Data: The dataset includes,

* Customers who left within the last month – the column is called Churn
* Services that each customer has signed up for – phone, multiple lines, internet, online security, online backup, device protection, tech support, and streaming TV and movies
* Customer account information – how long they’ve been a customer, contract, payment method, paperless billing, monthly charges, and total charges
* Demographic info about customers – gender, age range, and if they have partners and dependents

**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:**

By employing data mining techniques like logistic regression and decision trees, the study proposes a model to predict customers at risk of leaving. Implemented in R programming, the system allows for versatile dataset analysis. The goal is to help telecom companies proactively retain customers, thereby minimizing revenue loss associated with churn.

**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 telecommunications.

**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 Decision trees and logistic regression, 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 logistic regression.

- Experiment with feature engineering techniques to enhance model performance.

- Consider incorporating external data sources (e.g., market trends, credit-score 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 for potential enhancements.

**Conclusion**

In this document, we've outlined our approach to solving the problem of customer churn prediction using Decision trees and logistic regression. 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. By following this structured approach, we aim to create a reliable tool that contributes positively to the telecom industry.