**Customer Churn prediction**

**Phase 1**

**Annai college of Engineering and Technology**

**Submitted by:**

**Arun.B**

**Salmankhan.A**

**Imthiyas.k**

**Heram.P**

**Ahamed yasin.J**

**Problem Definition and Design Thinking**

Agenta:

1.Introduction

2.objectives

3.project definition

4.project scope

5.Data pipeline

6.Technology stack

7. Risks and Mitigation

8. Conclusion

**1. Introduction:**

The Customer Churn Prediction Project aims to develop a predictive model that identifies customers likely to churn. Churn prediction is crucial for businesses to retain their customer base and improve overall customer satisfaction. This document outlines the project scope, objectives, architecture, and implementation details.

**2. Objectives:**

- Develop an accurate machine learning model to predict customer churn.

- Evaluate and select the most suitable algorithms for churn prediction.

- Implement a user-friendly interface for accessing and utilizing the churn prediction model.

**3. Project Definition:**

 The project involves using IBM Cognos to predict customer churn and identify factors influencing customer retention. The goal is to help businesses reduce customer attrition by understanding the patterns and reasons behind customers leaving. This project includes defining analysis objectives, collecting customer data, designing relevant visualizations in IBM Cognos, and building a predictive model.

**4. Project Scope:**

**4.1 Data Collection**

- Data will be collected from various sources, including transaction records, customer interactions, and demographic information.

**4.2 Data Preprocessing**

- Data will be cleaned, normalized, and transformed for model compatibility.

- Feature selection and engineering will be performed to extract relevant information.

**4.3 Model Development**

- Multiple machine learning algorithms will be evaluated and compared for their churn prediction performance.

- Ensemble methods like Random Forest, Gradient Boosting, and Neural Networks will be considered.

**4.4 Model Evaluation:**

Performance metrics will include accuracy, precision, recall, F1-score, and AUC-ROC.

Cross-validation and hyperparameter tuning will be used to optimize the model.

**4.5 Model Deployment**

The selected model will be deployed in a production environment, allowing real-time predictions.

Containerization (e.g., Docker) will be used for easy deployment and scalability.

**4.6 User Interface**

Develop a user-friendly web interface for accessing the churn prediction model.

The interface will allow users to input customer information and receive churn predictions.

**5. Architecture**

**5.1 Data Pipeline**

Data will flow through a pipeline for collection, preprocessing, and feature engineering.

Tools such as Apache Airflow will be employed to automate and monitor the pipeline.

**5.2 Model Training and Evaluation**

Model training will be performed on a designated server using parallel processing for efficiency.

Evaluation will be conducted using a separate validation set and cross-validation.

**5.3 Model Deployment**

Docker containers will be used to package the model for deployment.

A web server (e.g., Flask) will serve as the interface for making predictions.

**5.4 Database**

A database (e.g., PostgreSQL) will store historical customer data and predictions for future analysis.

**6. Technology Stack:**

Programming Languages: **Python, SQL**

Machine Learning Libraries: **scikit-learn, TensorFlow, Kera's**

Web Framework: **Flask**

Database: **PostgreSQL**

Containerization: **Docker**

Workflow Automation: **Apache Airflow**

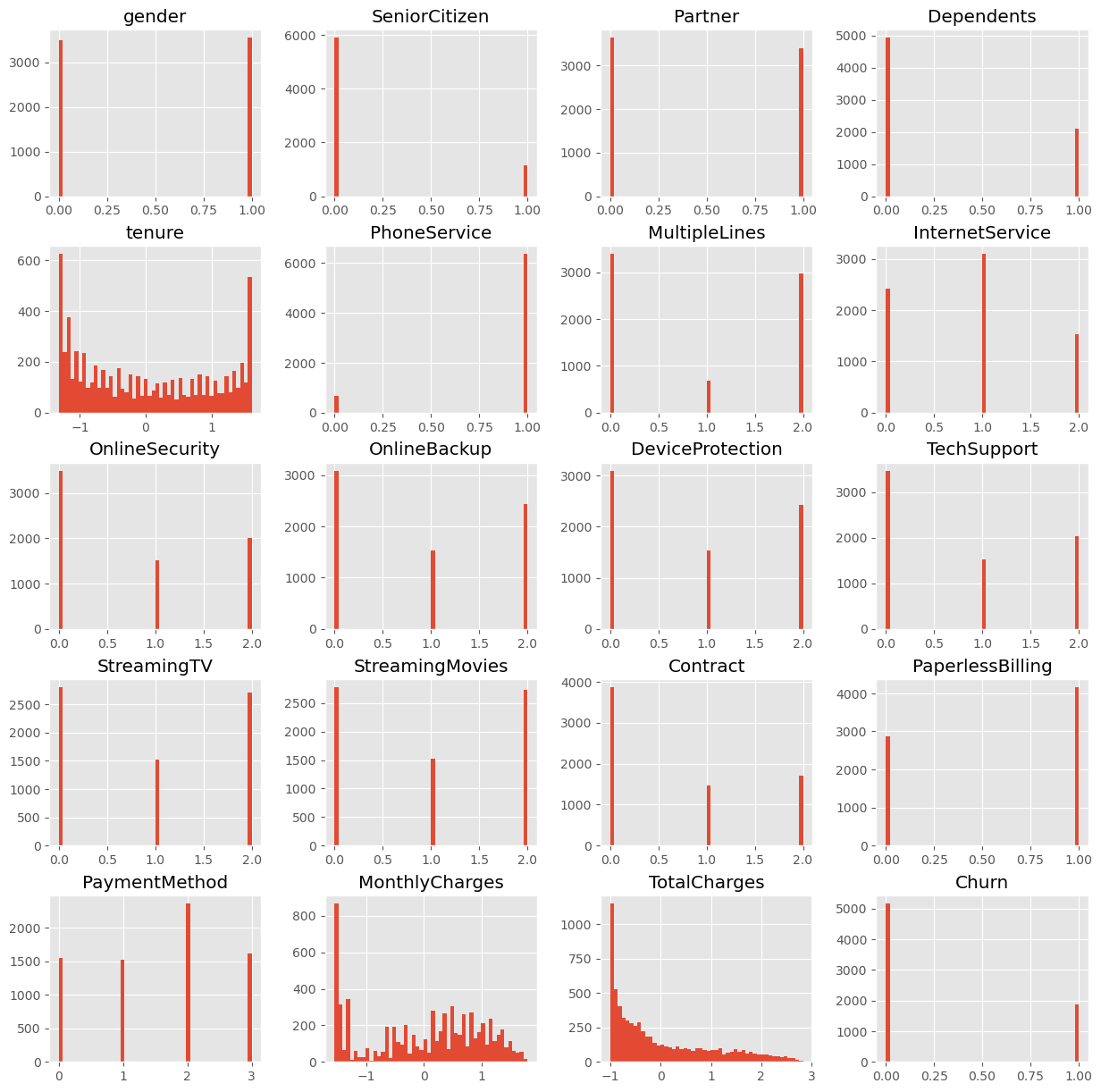
**7. Risks and Mitigation**

**Data Quality**: Conduct thorough data quality checks and implement robust preprocessing techniques.

**Model Overfitting:** Regularize models, perform cross-validation, and monitor performance on unseen data.

Scalability Issues: Design the system with scalability in mind, and conduct load testing.

**Visualization :**



**8. Conclusion**

The Customer Churn Prediction Project aims to deliver an accurate and reliable model for identifying potential churners. By implementing this system, businesses will be equipped with a tool to proactively address customer retention, leading to increased customer satisfaction and revenue.