Customer churn prediction

Nivetha S K
962221106079

Phase 1: Document submission
Project: **Customer churn prediction**

Phase 1: Problem definition and design thinking

Problem Statement: Predicting Customer Churn

Description: Customer churn, or customer attrition, refers to the rate at which customers stop doing business with a company. Predicting customer churn is crucial for businesses as it allows them to take proactive measures to retain valuable customers. The problem involves using historical customer data and various machine learning techniques to predict which customers are likely to churn in the future.

Data Collection: Gather historical customer data, including demographics, transaction history, customer interactions, and any relevant features.

Data Preprocessing: Clean and preprocess the data by handling missing values, encoding categorical variables, and scaling numerical features.

Prototype:

Create a prototype of the churn prediction system.

Design mockups of the user interface if applicable.

Develop a data pipeline for collecting and processing customer data.

Output: The output of the customer churn prediction model is a list of customers ranked by their likelihood of churning. This allows businesses to focus their retention efforts on high-risk customers and reduce customer attrition, ultimately improving customer satisfaction and profitability.

ANALYSIS APPROACH

The primary objectives of the customer churn prediction analysis approach are:

Reduce Churn Rates: The central goal is to identify customers who are likely to churn in the future. By doing so, a company can take proactive measures to retain those customers, thereby reducing overall churn rates and maintaining a healthy customer base.

Increase Customer Retention: The analysis aims to provide insights into why customers churn. By understanding the underlying reasons, a business can implement targeted retention strategies to address specific issues and improve customer satisfaction.

Optimize Marketing and Customer Engagement: Identifying which factors influence churn helps in tailoring marketing and customer engagement efforts. This ensures that resources are allocated effectively to retain high-risk customers and nurture loyal ones.

Improve Customer Experience: By analyzing customer data, a company can uncover pain points in the customer journey, enabling improvements in products, services, and support to enhance the overall customer experience.

Enhance Decision-Making: The analysis approach provides data-driven insights that aid in strategic decision-making. It helps in allocating resources more efficiently, prioritizing customer segments, and making informed decisions about pricing, promotions, and customer service initiatives.

Increase Revenue: Retaining existing customers is often more cost-effective than acquiring new ones. Predictive churn analysis can help in maintaining a stable customer base, leading to higher long-term revenue and profitability

DATA COLLECTION

Customer data can be collected from various sources using different methods. Here are common sources and methods for collecting customer data:

Database Queries: Extract relevant data from your company’s databases, CRM systems, and other data repositories.

Web Scraping: Gather data from websites, forums, and social media platforms where customers discuss your product or service.

API Integration: If available, use APIs to access data from third-party sources or platforms like social media.

Surveys and Questionnaires: Create and distribute surveys to collect customer feedback and preferences.

Call Records: Analyze call logs and recordings from customer support or sales calls.

Text Analytics: Apply natural language processing (NLP) techniques to analyze text data from emails, chat logs, and customer reviews.

IoT Devices: For businesses with IoT products, collect data from connected devices to track usage patterns.

Mobile App Analytics: If applicable, use analytics tools to gather data on mobile app usage

VIRTUALIZATION ANALYSIS

Implementing a virtualization strategy for customer churn prediction involves creating a virtual environment to simulate customer behavior and test predictive models. Here are steps to consider:

Data Collection and Preparation:

Gather historical customer data including demographics, transaction history, and interaction logs.

Clean and preprocess the data to remove outliers and missing values.

Virtual Environment Setup:

Create a virtual environment that mimics your real customer interactions.

Generate synthetic data or use historical data to simulate customer activities.

Feature Engineering:

Extract relevant features from the virtual environment, such as customer activity, engagement metrics, and customer attributes.

Model Development:

Develop machine learning models for churn prediction using the virtual data.

Train and optimize these models with synthetic data to simulate real-world scenarios.

Validation and Testing:

Use the virtual environment to validate and test the predictive models.

Measure the model’s performance in terms of accuracy, precision, recall, and F1-score.

Feedback Loop:

Continuously update and refine your models based on feedback from the virtual environment.

Adjust features, algorithms, or thresholds as needed.

Deployment in the Real World:

Once satisfied with model performance in the virtual environment, deploy the model to predict customer churn in your actual customer base.

Monitoring and Maintenance:

Monitor model performance in the real world and make adjustments as necessary.

Collect new data and periodically retrain the model to keep it up-to-date.

Scalability and Flexibility:

Ensure your virtualization strategy is scalable to accommodate larger datasets and changing customer behaviors.

Be prepared to adapt to evolving customer trends and preferences.

Ethical Considerations:

Be mindful of privacy and ethical considerations when working with customer data, even in a virtual environment.

Import necessary libraries

Import pandas as pd

Import numpy as np

```
From sklearn.model_selection import train_test_split

From sklearn.preprocessing import StandardScaler

From sklearn.ensemble import RandomForestClassifier

From sklearn.metrics import accuracy_score, confusion_matrix, classification_report

Import matplotlib.pyplot as plt

# Load your dataset

Data = pd.read_csv(“customer_data.csv”)  # Replace with your dataset file

# Data preprocessing

# - Handle missing values

# - Encode categorical features (e.g., one-hot encoding)

# - Scale numerical features

# - Define the target variable (e.g., ‘Churn’) and features

# Split the data into training and testing sets

X_train, X_test, y_train, y_test = train_test_split(features, target, test_size=0.2,
random_state=42)

# Feature scaling

Scaler = StandardScaler()

X_train = scaler.fit_transform(X_train)

X_test = scaler.transform(X_test)

# Initialize and train a machine learning model (e.g., Random Forest)

Model = RandomForestClassifier(n_estimators=100, random_state=42)

Model.fit(X_train, y_train)
```

```
# Make predictions

Y_pred = model.predict(X_test)

# Evaluate the model

Accuracy = accuracy_score(y_test, y_pred)

Conf_matrix = confusion_matrix(y_test, y_pred)

Report = classification_report(y_test, y_pred)

# Display results

Print(f'Accuracy: {accuracy}")

Print(f'Confusion Matrix:\n{conf_matrix}")

Print(f'Classification Report:\n{report}")

# Visualize important features

Feature_importances = model.feature_importances_

Plt.barh(features.columns, feature_importances)

Plt.xlabel('Feature Importance')

Plt.ylabel('Features')

Plt.show().
```

CONCLUSION

In conclusion, Phase 1 of our churn prediction analysis project lays the foundation for our future efforts. We have successfully collected and preprocessed the necessary data, identified key features, and performed initial exploratory data analysis.