Project 6: Customer Churn Prediction

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.

Design thinking

1. Analysis Objectives: Define the specific objectives of predicting customer churn, such as identifying potential churners and understanding the key factors contributing to churn.

2. Data Collection: Determine the sources and methods for collecting customer data, including customer demographics, usage behavior, and historical interactions.

3. Visualization Strategy: Plan how to visualize the insights using IBM Cognos, showcasing factors affecting churn and retention rates.

4. Predictive Modeling: Decide on the machine learning algorithms and features to use for predicting customer churn.

Analysis Objective

1: Identify Potential Churners

Identifying customers who are likely to churn is the primary objective. By analyzing historical data and customer behavior patterns, businesses aim to accurately classify customers into potential churners and non-churners.

2: Understand Churn Patterns and Trends

The analysis should aim to identify patterns and trends in customer churn. This involves studying historical churn data over different time periods to recognize fluctuations and seasonal trends. Understanding when churn rates are typically high or low provides insights for timely interventions and resource allocation.

3: Determine Key Churn Drivers

Determining the factors leading to churn is crucial. By analyzing customer data, feedback, and interactions, businesses can identify specific factors such as poor customer service, pricing issues, or product dissatisfaction. Pinpointing these drivers is essential for devising strategies that directly address the root causes of churn.

4: Develop Predictive Models

Developing predictive models using machine learning algorithms is a significant objective. These models utilize historical data and identified churn drivers to forecast future churn. Accurate predictive models enable businesses to proactively identify potential churners, allowing for timely and personalized retention efforts.

5: Segment Customers for Targeted Strategies

Segmentation of the customer base based on behavior, demographics, or other relevant factors is essential. Analyzing churn within these segments helps in tailoring retention strategi

Data collection

1. **Customer Surveys and Feedback:**

- **Sources:** Conducting online or in-person surveys, email questionnaires, and feedback forms on websites or mobile apps.
- **Methods:** Designing well-structured surveys with questions about demographics, preferences, satisfaction levels, and reasons for using or leaving the service.

2. **Website and App Analytics:**

- **Sources:** Website analytics tools (like Google Analytics) and in-app analytics solutions.
- **Methods:** Tracking user behavior, such as pages visited, time spent on site,

3. **Choose Appropriate Visualizations:**

- **Churn and Retention Rates:** Use line charts or bar charts to show trends over time. Line charts are great for showing changes over continuous time periods, while bar charts can effectively compare values for different categories (months, quarters, etc.).
- **Factors Affecting Churn:** Utilize bar charts, heat maps, or pie charts to represent categorical data. Bar charts can show the frequency of different factors, heat maps can provide a visual representation of correlations, and pie charts can represent the proportion of each factor contributing to churn.

4. **Customer Support Interactions:**

- **Sources:** Customer support emails, chats, phone calls, and helpdesk tickets.
- **Methods:** Analyzing customer queries, complaints, and feedback. Customer relationship management (CRM) systems can be used to store and analyze this data.

5. Transaction and Purchase History:

- **Sources:** Sales databases, e-commerce platforms, and point-of-sale systems.
- **Methods:** Recording purchase history, including products or services bought, purchase frequency, transaction amount, and payment methods. Analyzing this data provides insights into customer preferences and buying behavior.

Visualization

1. Define the Audience:

Identify the stakeholders and their specific needs. Understand the level of detail they require and the key metrics they are interested in, such as churn rates, retention rates, or factors influencing these rates.

2. Data Preparation:

Ensure that your data is clean, complete, and well-organized. Use IBM Cognos tools to connect to your data sources, perform necessary transformations, and create a consolidated dataset that includes information about customer demographics, usage behavior, historical interactions, churn status, and retention efforts.

3. Choose Appropriate Visualizations:

- **Churn and Retention Rates:** Use line charts or bar charts to show trends over time. Line charts are great for showing changes over continuous time periods, while bar charts can effectively compare values for different categories (months, quarters, etc.).
- **Factors Affecting Churn:** Utilize bar charts, heat maps, or pie charts to represent categorical data. Bar charts can show the frequency of different factors, heat maps can provide a visual representation of correlations, and pie charts can represent the proportion of each factor contributing to churn.

4. Interactive Dashboards:

Create interactive dashboards in IBM Cognos to allow stakeholders to explore the data. Include filters and parameters that enable users to drill down into specific segments, time periods, or other relevant dimensions. Interactive elements like drop-down menus and sliders enhance user engagement.

5. Storytelling Approach:

Construct a narrative around your visualizations. Arrange visualizations in a logical sequence to tell a compelling story. Start with an overview of churn and retention rates, then delve into specific factors affecting churn. Use annotations and text boxes to highlight key observations and insights.

Predictive Modeling:

1. Understanding the Data:

- **Data Exploration:** Analyze the dataset to understand its characteristics, including the types of features available, data distributions, and potential correlations.

- **Feature Importance:** Identify key features that are likely to influence customer churn. This could include customer demographics, transaction history, customer support interactions, and usage patterns.

2. **Choosing Machine Learning Algorithms:**

- **Binary Classification Algorithms:** Customer churn prediction is typically framed as a binary classification problem, where customers are classified as churners or non-churners.
- **Common Algorithms:** Consider algorithms such as Logistic Regression, Decision Trees, Random Forest, Gradient Boosting (e.g., XGBoost, LightGBM), and Neural Networks (for more complex relationships). These algorithms often perform well for churn prediction tasks.
- **Ensemble Methods:** Ensemble methods like Random Forest and Gradient Boosting can improve prediction accuracy by combining predictions from multiple models.
- **Deep Learning:** For large, complex datasets with intricate patterns, deep learning models like Artificial Neural Networks (ANNs) can be explored.

3. **Feature Selection and Engineering**:

- **Feature Selection:** Use techniques like correlation analysis and feature importance scores from tree-based models to select the most relevant features.
- **Feature Engineering:** Create new features from existing ones that might carry valuable information. For example, you could derive metrics like customer tenure, transaction frequency, or customer satisfaction scores from the available data.

4. **Handling Imbalanced Data:**

- **Imbalanced Data:** Churn datasets often have imbalanced classes, where the number of churners is significantly lower than non-churners.
- **Addressing Imbalance:** Techniques like oversampling the minority class (churners), undersampling the majority class, or using Synthetic Minority Over-sampling Technique (SMOTE) can be employed to balance the dataset and improve model performance.

5. **Validation and Evaluation Metrics:**

- **Validation Split:** Split the dataset into training and validation sets. Use techniques like cross-validation to robustly evaluate model performance.
- **Evaluation Metrics:** Besides accuracy, consider metrics like precision, recall, F1-score, and Area Under the ROC Curve (AUC-ROC). Precision focuses on the accuracy of positive predictions, while recall assesses the ability to capture all positives.

Teammates

 Arun pragash M

Arockia mathan S

Parimaleshwaran A

Adharsh R
Arun kumaran P