Project 6:

# Customer Churn Prediction

**SUBMITTED BY :**

**BALISETTY CHENCHU NITHIN**

**Au723921104003**

**bcnnithin@gmail.com**

Project Title: Customer Churn Prediction

Project Steps

Phase 1: Project Definition and Design Thinking

**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.

The specific objectives of predicting customer churn include identifying potential churners and understanding the key factors contributing to churn. By analyzing customer data and behavior patterns, the aim is to accurately predict which customers are likely to churn in the future. This information can help businesses take proactive measures to retain customers and address the underlying issues that lead to churn.

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

1. Customer Demographics: Customer demographic information can be collected through surveys, registration forms, or customer profiles. This may include age, gender, location, occupation, etc.

2. Usage Behavior: Usage behavior data can be collected through tracking customer interactions with the product or service. This can include data on frequency of usage, duration of usage, features utilized, and any patterns or trends observed.

3. Historical Interactions: Historical interaction data can be collected from customer support logs, call center records, chat transcripts, or email communications. This data can provide insights into past interactions, complaints, inquiries, or feedback from customers.

4. Website and App Analytics: Website and app analytics tools can provide valuable data on customer behavior, such as page views, click-through rates, bounce rates, and conversion rates. This data can help understand how customers engage with the digital platforms.

5. Social Media Monitoring: Monitoring social media platforms can provide insights into customer sentiment, brand mentions, and customer feedback. This can be done through social listening tools or manual monitoring of relevant social media channels.

It is important to ensure that data collection methods comply with privacy regulations and obtain necessary consent from customers. Additionally, data security measures should be implemented to protect customer information.

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

To visualize insights about factors affecting churn and retention rates using IBM Cognos, follow these steps:

1. **Data Preparation**:

- Ensure you have clean and well-structured data that includes information about customer churn and relevant factors like demographics, purchase history, customer interactions, etc.

2. **Connect to Data Source**:

- Open IBM Cognos and connect to your data source, which could be a database, spreadsheet, or other data storage.

3. **Create Data Modules**:

- Build data modules in IBM Cognos to define relationships between different data tables or sources. Ensure that you have a clear understanding of how customer information relates to churn and retention.

4. **Create Queries**:

- Construct queries to extract the necessary data for your analysis. You'll want to select data related to customer churn and relevant factors like customer demographics, behavior, and interactions.

5. **Data Exploration**:

- Use Cognos' data exploration tools to get an initial understanding of your data. Identify patterns, trends, and outliers that could be affecting churn and retention rates.

6. **Build Visualizations**:

- Create visualizations like charts, graphs, and tables to represent your insights. Common types of visualizations for churn and retention analysis include:

  - Line charts showing trends over time.

  - Pie or bar charts for demographic breakdowns.

  - Heatmaps for correlation analysis.

  - Cohort analysis charts to track customer cohorts over time.

7. **Segmentation**:

- Segment your customer data to analyze different groups separately. For example, segment by age, location, or product usage. This will help you identify which segments are more prone to churn.

8. **Calculate Churn and Retention Rates**:

- Use Cognos to calculate churn and retention rates based on your data. Churn rate is typically calculated as the number of customers who churned divided by the total number of customers, while retention rate is the inverse (1 – churn rate).

9. **Create Dashboards**:

- Design interactive dashboards that combine your visualizations and key metrics. Dashboards provide a high-level overview of the insights and can be shared with stakeholders.

10. **Add Filters and Interactivity**:

- Allow users to filter data and explore insights interactively. This enables users to drill down into specific segments or time periods for deeper analysis.

11. **Schedule and Share Reports**:

- Schedule regular reports or updates to keep stakeholders informed. You can also share reports through email or a Cognos portal for wider distribution.

12. **Continuously Monitor**:

- Churn and retention rates change over time, so it's essential to set up regular monitoring and update your visualizations and analyses as new data becomes available.

Remember that IBM Cognos provides a wide range of tools and functionalities for data visualization and analysis. Customizing your approach to suit your specific data and business needs is key to effectively showcasing factors affecting churn and retention rates.

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

Predicting customer churn is a critical task, and selecting the right machine learning algorithms and features is crucial for accurate predictions. Here's a general framework to help you decide:

**Machine Learning Algorithms**:

1. **Logistic Regression**:

- It's a good starting point for binary classification problems like churn prediction.
- Provides interpretable results and is easy to understand.

2. **Random Forest**:

- Effective for handling complex relationships in data and capturing feature importance.
- Robust to outliers and missing data.

3. **Gradient Boosting (e.g., XGBoost, LightGBM)**:

- Powerful ensemble methods that often perform well in churn prediction tasks.
- Tends to handle imbalanced datasets well.

4. **Neural Networks**:

- Deep learning models like feedforward neural networks can capture intricate patterns in the data.

- Require more data and computational resources but can offer high predictive accuracy.

5. **Support Vector Machines (SVM)**:

- Useful when dealing with high-dimensional data and non-linear decision boundaries.

6. **K-Nearest Neighbors (KNN)**:

- Simple and effective for smaller datasets, especially when there's spatial clustering.

7. **Naive Bayes**:

- Fast and efficient for text data or when you have limited computational resources.

8. **Time-Series Models (e.g., ARIMA, LSTM)**:

- If your churn data has a strong temporal component, time-series models may be appropriate.

**Feature Selection and Engineering**:

1. **Feature Selection**:

- Use techniques like feature importance from tree-based models or recursive feature elimination to identify the most relevant features.

2. **Feature Engineering**:

- Create new features that might have predictive power. For churn prediction, consider:

  - Customer demographics (age, gender, location).

  - Customer behavior (purchase history, frequency, recency).

  - Customer interactions (customer support calls, complaints).

  - Usage patterns (app or website engagement).

3. **One-Hot Encoding**:

   - Convert categorical variables into numerical form using one-hot encoding.

4. **Scaling and Normalization**:

   - Ensure that numerical features are scaled or normalized to prevent one feature from dominating others.

5. **Handling Imbalanced Data**:

   - If you have imbalanced classes (e.g., fewer churn cases), consider techniques like oversampling, undersampling, or using class weights during training.

6. **Feature Crosses**:

   - Create interaction features to capture combined effects of two or more features.

7. **Text Data Processing** (if applicable):

   - If you have textual data (e.g., customer feedback), use natural language processing techniques for sentiment analysis or topic modeling.

8. **Time-Based Features** (if applicable):

   - If your data is time-series, create features that capture trends, seasonality, and temporal patterns.

Remember to split your data into training and testing sets to evaluate the performance of different algorithms and features. Experiment with different combinations to find the model that provides the best predictive accuracy for your specific churn prediction problem. Additionally, consider using techniques like cross-validation for robust model evaluation.