Customer Churn Prevention Recommender System

1) What is the purpose of customer churn prediction?

The model's primary purpose in a customer churn predictor is to identify customers likely to stop using a product or service (i.e., "churn") within a specific timeframe. By knowing this data, businesses can take proactive steps, such as targeted retention campaigns or offering special incentives, to prevent these customers from leaving.

Customer churn refers to losing customers or clients, typically measured as the percentage of customers who discontinue their relationship with a business during a given period.

We create a churn prediction model using **Catboost classifier** after comparing different types of algorithms on our data we choose the best algorithm.

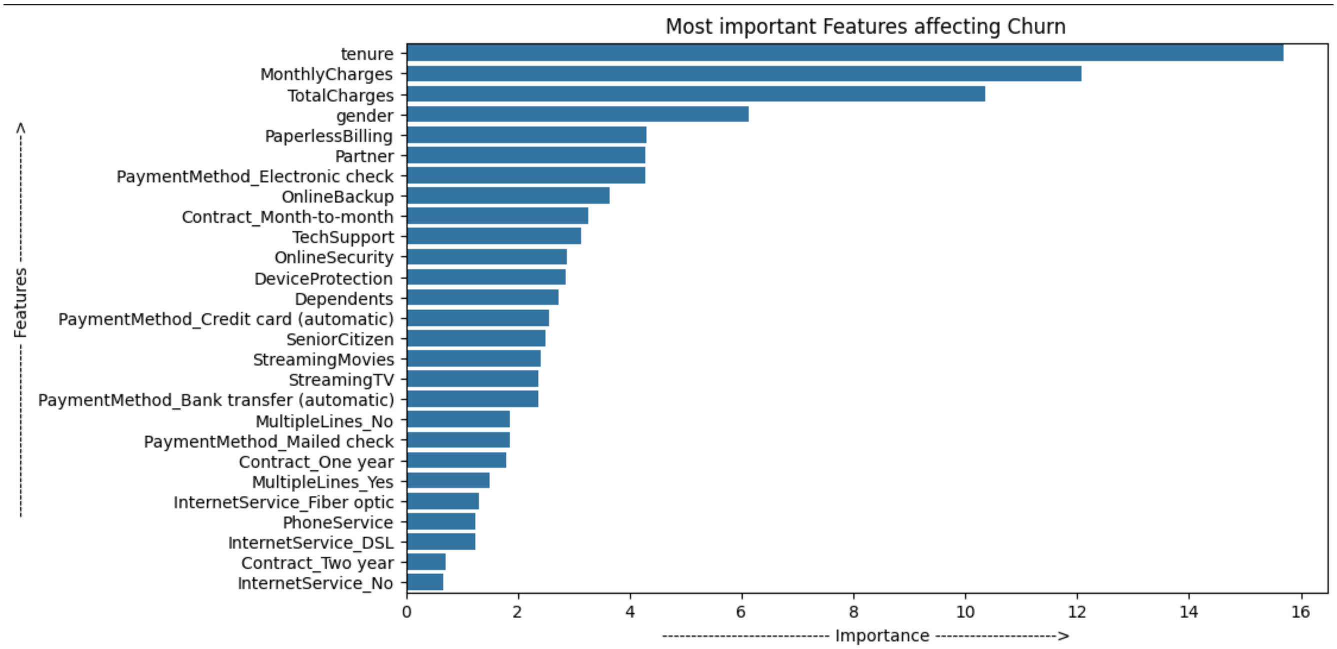
The **insights from the model are then used in our Recommender system** to give personalized recommendations to customers based on the services they are using currently.

2) Dataset used

- **Relevance to Churn Prediction**: Includes key customer data like tenure, contract type, charges, and services - ideal for analyzing churn and retention strategies.
- **Diverse Features**: Contains numerical (tenure, charges) and categorical (gender, payment method) data, allowing for rich analysis across multiple customer dimensions.
- **Real-World Applicability**: Reflects a genuine telco business context, making insights and recommendations practical for similar industries which are customer-oriented.
- **Substantial Data Size**: Enough records to enable meaningful pattern discovery and model training while remaining computationally manageable.
- **Clearly Labeled Target**: The "Churn" column simplifies supervised learning, directly supporting classification models.
- **Balanced Complexity**: Offers an ideal level of detail for showcasing machine learning applications without being overwhelming, making it valuable for end-to-end projects.

The data set includes information about :

- Customers who left within the last month, this is the output column, 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
- Tenure of services and type of subscription - monthly / yearly / etc.
- 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
- Gives a wide range of information about the customer which will help us in understanding the mindset of users and accordingly increase the revenue of the company by selling more plans to the consumers.



3) Churn prediction model

The comparison of model accuracies indicates how well each algorithm performed on the churn prediction task:

- Logistic Regression (Accuracy = 0.8049): Performs well for this type of binary classification and provides a solid baseline model due to its simplicity and interpretability.
- K-Nearest Neighbors (KNN) (Accuracy = 0.7656): Lower accuracy suggests that it may struggle with complex relationships in the data and sensitivity to feature scaling.

- Decision Tree (Accuracy = 0.7234): While interpretable, it may have overfitted the training data, resulting in lower accuracy on unseen data.
- Random Forest (Accuracy = 0.7938): An improvement over a single decision tree due to ensemble learning, but still needs to be higher than other boosting methods.
- Support Vector Machine (SVM) (Accuracy = 0.7985): Provides good performance but may require careful tuning of hyperparameters and kernel selection.
- Gradient Boosting (Accuracy = 0.8028): Strong performance is a boosting method that iteratively improves weak learners but is slightly lower than CatBoost.
- CatBoost (Accuracy = 0.86): Outperformed other models due to its superior handling of categorical features, overfitting prevention, and effective learning of complex relationships without extensive preprocessing.

4) Recommendation models :

What is a recommender system?

A **recommender or recommendation system** is a subclass of information filtering systems that provides suggestions for items most pertinent to a particular user. Widely used nowadays in all services based companies where the main focus of concern is consumer in such cases recommendation systems have crucial importance.

Types of Recommender Systems:

1. **Content-Based Filtering**:
    - This approach recommends items based on the items' attributes and the user's preferences. It relies on the customers' features and recommends services similar to those they have used.
2. **Collaborative Filtering**:
    - This method suggests items based on the preferences and behaviours of similar users. It requires a larger dataset to identify user patterns, which is beneficial for leveraging insights from a community of customers.
3. **Hybrid Systems**:
    - Combining content-based and collaborative filtering techniques, hybrid systems aim to enhance recommendation accuracy by incorporating multiple data sources.

Choice of Recommender System:

For this project, we are utilizing a **collaborative-based filtering** approach for the following reasons:

- **Feature Richness**: The model leverages various customer features, allowing for precise recommendations tailored to individual customer profiles based on their unique attributes and service usage.
- **Data Availability**: Given the structured nature of our dataset, collaborative-based filtering effectively utilizes the attributes of each customer to identify similar customer profiles and suggest appropriate services.
- **Interpretability**: Collaborative-based recommendations are much more accurate, as they can directly relate to the features and services they are using or considering and most probably will have similar thought of action process.
- **Cold Start Problem Mitigation**: This approach can provide meaningful recommendations even for new customers by relying on their immediate input features without needing extensive historical data on other users.

Input Handling / Data Processing :

- **Feature Collection**: The model accepts various customer attributes as input:
  - **Demographics**: Gender (categorical), Senior Citizen (binary), Partner (binary), Dependents (binary).
  - **Account Metrics**: Tenure (numeric), Monthly Charges (numeric), Total Charges (numeric), Paperless Billing (binary).
  - **Services & Subscriptions**: Phone Service (binary), Multiple Lines (binary), Internet Service (definite), Online Security (binary), Online Backup (binary), Device Protection (binary), Tech Support (binary), Streaming TV (binary), Streaming Movies (binary).
  - **Contract Type**: (categorical), Payment Method (categorical).
- **Encoding**:
  - Categorical features undergo one-hot encoding to convert into a numerical format suitable for machine learning models providing better performance of the models. For instance, `Gender` is transformed into two binary features:
  - (`Gender_Male` and `Gender_Female which will have value 0 or 1 )`
  - This results in a data frame that aligns with the feature space of the trained model.
- **Feature engineering :**
  - Dropping unnecessary features from data which doesn’t contribute to final output, this helps in reducing the dimensions of data improving computation time.
  - Standardize the input using techniques like standard scaler and normalization.
  - Removing imbalance from the data through Oversampling by performing techniques like SMOTE (Synthetic Minority Oversampling Technique).

Feature Dependencies:

- **Demographics**:
  - **Gender**: Provides insights into potential service preferences and engagement strategies based on the gender of the consumer.
  - **Senior Citizen Status**: Influences the need for additional customer support services and product offerings tailored for seniors catering to their needs.
- **Account Metrics**:
  - **Tenure**: Longer tenure generally correlates with customer loyalty; a direct relationship with churn likelihood informs retention strategies.
  - **Monthly Charges & Total Charges**: Analyzing these charges helps identify customers who may feel they are not receiving sufficient value, indicating a need for tailored retention offers.
- **Service Subscriptions**:
  - Features like **Online Security** and **Tech Support** indicate service gaps. The lack of these services in a customer’s profile may prompt recommendations for enhancements based on patterns observed in similar customers who have avoided churn.
  - **Internet Service Type**: Affects satisfaction levels and churn risk. Recommendations suggest upgrades if a customer uses lower-tier services (e.g., DSL) compared to higher-tier services (e.g., Fiber optic).

Similarity Calculation:

- The model employs Cosine Similarity to compute the similarity matrix among customer profiles, using the input features as vectors to dfinmd the similarity among the customers with similar behavior.
- For a target customer, the model retrieves the top “N” similar customers based on their feature vectors, which helps identify peers with similar characteristics and their corresponding behaviors regarding churn.

Recommendation Generation:

- The model uses similar customers identified to analyze their service usage and churn outcomes deriving common qualities.
- It generates tailored recommendations by:
    - Assessing average usage patterns of contract types and additional services among non-churned similar customers.
    - It proposes popular services or contract types among peers who have shown a lower churn rate.
- For instance, if similar customers with long-term contracts have higher retention rates, the model may recommend switching to a long-term contract for the target customer on a month-to-month plan.

5) Results :

- **Informed Decision-Making**: Enables targeted retention strategies and efficient resource allocation.
- **Personalized Customer Experience**: Offers customized tailored recommendations, enhancing satisfaction and engagement.
- **Enhanced Customer Retention**: Facilitates proactive interventions to retain at-risk customers and build long-term relationships through better retention.
- **Business Insights**: Provides valuable insights into customer behavior, guiding product and service refinement.
- **Performance Measurement**: Allows tracking of churn rates and effectiveness of retention strategies for continuous improvement.
- **Competitive Advantage**: Utilizes data-driven insights to stay ahead of competitors and rivals while also meeting customer needs and improvised facilities.