**University of Petroleum and Energy Studies**

**Internship - Low Level Design**

**on**

**Predicting Customer Churn in Telecommunication Company**

Team members:

Jitender Dhariwal - R2142201723

Jatin Wadhwa - R214220562

Kartikay Tyagi -R214220603

Jatin Joshi - R214220559

Karan Taneja - R214220587

Guided by: Mr. Sumit Shukla

Industry Mentor: Mr. Sumit Shukla

**Table of Contents**

1. Introduction
   1. Scope of the document
   2. Intended audience
   3. System overview
2. Low Level System Design
   1. Sequence Diagram
   2. Navigation Flow/UI Implementation
   3. Components Design Implementation
   4. Configurations/Settings
3. Data Design
   1. List of Key Schemas/Tables in database
   2. Details of access levels on key tables in scope
   3. Key design considerations in data design
4. Details of other frameworks being used.
5. Reference

**1. Introduction**

For telecommunications firms, client churn—the process in which consumers transfer from one provider to another—raises serious concerns. Retaining clients has risen at the top of many companies' priorities due to fierce competition and changing market conditions. Telecommunications businesses are increasingly using machine learning techniques to forecast client attrition in an effort to reduce the negative effects of churn.

Machine learning models provide an effective way to examine vast amounts of consumer data and spot trends that may be used to identify which customers are most likely to leave. These models can assist telecoms firms take preventative steps to retain at-risk consumers by making use of past customer information.

In the telecoms sector, the aim of customer churn prediction using machine learning is to create models that can precisely identify consumers who are likely to churn in the near future. Companies can enhance customer happiness and loyalty by recognizing these clients in advance and taking the required actions, such as making personalized retention offers, performing outstanding customer service, or launching focused marketing efforts.

Telecommunications firms are able to more efficiently manage their resources because to the predictive capabilities of machine learning algorithms. These models help businesses to concentrate their efforts on consumers who are most likely to churn, allowing them to optimize their retention tactics and lower the expenses involved with gaining new customers.

Furthermore, customer churn prediction models can provide insights into the factors that drive customer attrition. By analyzing the most important features contributing to churn, such as poor network quality, pricing dissatisfaction, or competitor offers, companies can gain a deeper understanding of customer preferences and pain points. This information can guide strategic decision-making, product development, and customer experience improvements.

In this era of data-driven decision-making, customer churn prediction using machine learning offers telecommunications companies a competitive advantage. By harnessing the power of advanced analytics, these companies can enhance customer retention efforts, increase customer satisfaction, and ultimately improve their bottom line.

In this guide, we will explore the steps involved in building a machine learning model for customer churn prediction in a telecommunications company. We will discuss data collection, pre-processing, feature engineering, model selection, training, evaluation, and deployment. By following these steps, telecommunications companies can develop robust churn prediction models and make data-driven decision.

**Scope of the document**

The different stakeholders within a telecom firm who can gain from the insights and forecasts produced by the churn prediction system make up the intended audience of customer churn prediction in the telecom industry. These parties play a role in business strategy, customer relationship management, and decision-making. Key target audiences include the following:

Marketing and sales teams can use churn forecasts to create targeted marketing campaigns and promotional offers to keep hold of high-value clients. Cross-selling and upselling opportunities can be found with the use of churn projections.

Customer Retention Teams: Customer retention teams are essential to the implementation of churn-reduction measures. Churn projections give them useful information about high-risk clients, enabling them to concentrate on tailored retention initiatives.

Customer support teams can use churn projections to proactively solve customer issues and deliver better service, ultimately increasing customer satisfaction and loyalty.

Product Development Teams: Churn forecasts can help product development teams to better understand customer preferences and pain areas so they can design services that better satisfy those requirements and expectations.

corporate analysts: To better understand customer behavior, spot trends, and offer data-driven advice to improve corporate performance, business analysts examine churn prediction data.

Executives and Management: In order to make strategic decisions, distribute resources effectively, and promote long-term business success, a telecom company's senior management depends on churn estimates.

**Intended audience**

The intended audience of customer churn prediction in the telecom industry includes various stakeholders within a telecom company who can benefit from the insights and predictions generated by the churn prediction system. These stakeholders are involved in decision-making, customer relationship management, and business strategy. Some of the key audience groups are:

Marketing and Sales Teams: Marketing and sales teams can leverage churn predictions to develop targeted marketing campaigns and promotional offers to retain high-value customers. Churn predictions can also help in identifying opportunities for cross-selling and upselling services.

Customer Retention Teams: Customer retention teams play a critical role in implementing strategies to reduce churn. Churn predictions provide them with valuable insights on high-risk customers, allowing them to focus on personalized retention efforts.

Customer Support and Service Teams: Customer support teams can utilize churn predictions to proactively address customer issues and provide better service, ultimately enhancing customer satisfaction and loyalty.

Product Development Teams: Churn predictions can offer insights into customer preferences and pain points, guiding product development teams to design services that better meet customer needs and expectations.

Business Analysts: Business analysts analyze churn prediction data to understand customer behavior, identify trends, and make data-driven recommendations to improve business performance.

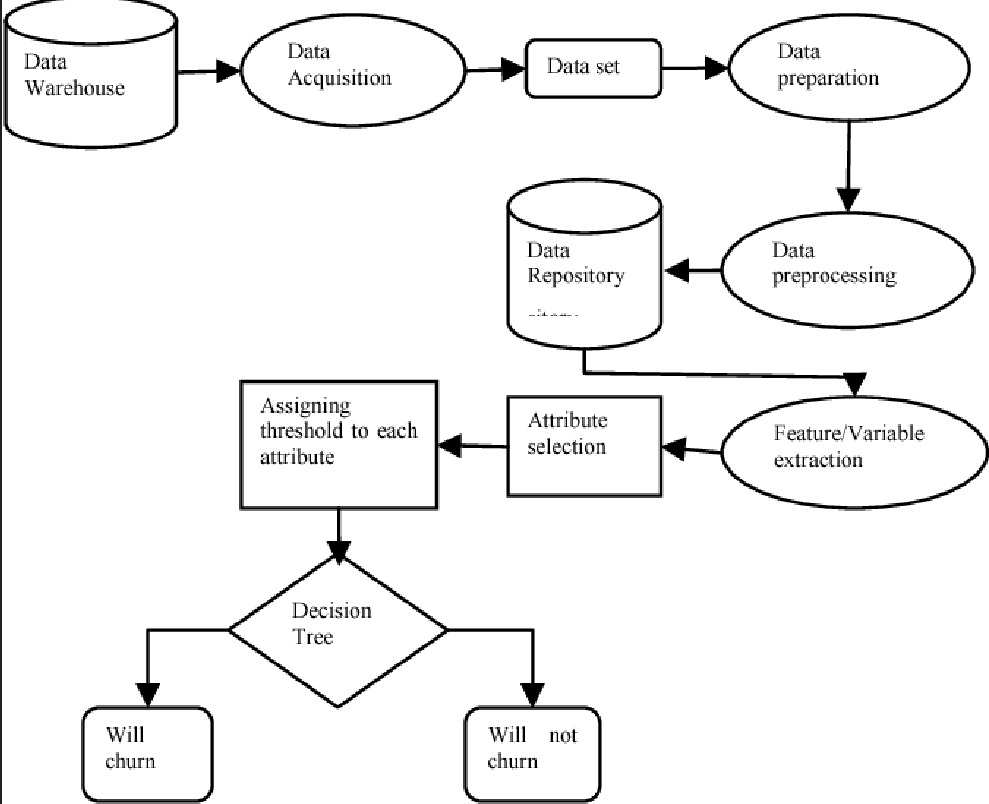
Executives and Management: The top management of a telecom company relies on churn predictions to make strategic decisions, allocate resources efficiently, and drive long-term business growth.

**System overview**

A telecommunications business can use the churn prediction model to identify customers who are likely to abandon their services in the near future. The organization can use targeted retention measures to lower customer churn and boost customer loyalty by proactively identifying potential churners. Obtaining historical client data from a variety of sources, including user profiles, call logs, payment details, usage statistics, and customer service interactions, is the first stage. In order to handle missing values, normalize the data, and engineer features, this data is then preprocessed. To build the model, pertinent features are chosen from the preprocessed data. These characteristics could include things like client demographics, usage habits, tenure, contract type, call length, complaints from customers, and more. In the selection process, qualities that are instructive are prioritized over those that are unnecessary or redundant. The chosen features are trained using past customer data by a machine learning model. For churn prediction, methods including logistic regression, decision trees, random forests, support vector machines (SVM), gradient boosting, and neural networks are frequently utilized. To properly assess the performance of the model, the dataset is divided into training and validation sets. Metrics including accuracy, precision, recall, F1-score, and ROC-AUC are used to assess the model's performance. To guarantee the model's generalizability and avoid overfitting, cross-validation techniques may be used. Following model training, a cutoff point is selected to categorize consumers as churners or non-churners based on the anticipated likelihood scores. According on the company's priorities, the threshold number can be changed to create a balance between false positives and false negatives. To process real-time customer data, the trained churn prediction model is incorporated into the telecom company's current systems. The model forecasts the likelihood of customer churn for specific customers as fresh data comes in. Customer segmentation can also be done using the churn prediction model. Based on customer turnover probability, customers are divided into various segments, which enables the business to more precisely focus retention initiatives. High-risk churners, for instance, can get unique offers to entice them to stay. The telecommunications business might execute particular retention strategies with the help of attrition projections and client groups. To decrease turnover and increase customer satisfaction, these techniques could include providing discounts, personalized offers, better customer service, loyalty programs, or focused marketing initiatives. The performance and accuracy of the model are constantly checked to guarantee its longevity. In order to adjust the model to changing client behavior and trends, new data or updated machine learning algorithms may be used.

**Low Level System Design**

**Sequence Diagram**



**Navigation Flow**:

The navigation flow of customer churn prediction in the telecom industry typically involves a series of steps and decision points to identify potential churners. Below is an outline of the navigation flow:

* Data Collection: Gather relevant data from various sources, such as customer profiles, call records, usage history, payment information, and customer service interactions.
* Data Preprocessing: Cleanse the data by handling missing values, outliers, and data inconsistencies. Transform and format the data into a suitable structure for analysis.
* Feature Engineering: Identify and select relevant features (predictors) that may impact customer churn, such as call duration, call frequency, contract type, data usage, complaints, etc.
* Splitting data: To train the model and assess how well it works, divide the dataset into training and testing sets.
* Model choice: For churn prediction, use relevant machine learning techniques like Logistic Regression, Decision Trees, Random Forest, Gradient Boosting, or Neural Networks.
* Model Education: Utilize the training dataset to run the chosen model.
* Model Evaluation: Evaluate the model's performance using the testing dataset.

Utilize metrics such as accuracy, precision, recall, F1-score, ROC-AUC, etc., to assess the model's predictive power.

* Model Tuning: Fine-tune model hyperparameters to improve performance if necessary.
* Customer Churn Prediction: Use the trained model to predict customer churn for new data. Calculate the churn probability for each customer based on model predictions.
* Customer Segmentation: Segment customers based on their churn probability or other relevant criteria.

**Components Design Implementation**

A machine learning (ML) project for reducing customer churn in a telecommunications company requires the design and implementation of several crucial elements. An overview of the key elements and Python's implementation is provided below:

* Data gathering and preparation: Collect past client information, such as demographics, usage trends, contract information, customer service encounters, etc.Handling missing values, encoding categorical variables, scaling numerical features, and dividing the dataset into training and testing sets are all examples of preprocessing.
* Engineering Features: Identify pertinent elements from the raw data that may help the model operate more predictably. As an illustration, consider computing customer tenure, averaging consumption data across time, developing binary indicators for certain services used, etc.
* Model choice: For the churn prediction task, pick suitable machine learning techniques like Logistic Regression, Random Forest, Gradient Boosting, or Neural Networks. Use libraries like scikit-learn, tensorflow, or keras to implement various models.
* Model Education: On the training dataset, run the selected models. To avoid overfitting, analyze and fine-tune model hyperparameters using methods like cross-validation.
* Model assessment: Utilize pertinent assessment metrics, including as accuracy, precision, recall, F1-score, ROC-AUC, etc., to evaluate how well the trained models performed on the test dataset.
* Observation and Logging: To keep track of incoming prediction requests and model performance over time, use logging systems.
* Updating and maintaining data: To preserve the model's predicted accuracy as new client data becomes available, schedule regular updates. Create a procedure for updating the model on a regular basis or if the distribution of the data significantly changes.

**Configurations/Settings**

1. Data Preprocessing Settings: Handling Missing Values: Specify the strategy for dealing with missing values in the dataset (e.g., imputation with mean, median, or mode).

Feature Scaling: Decide on the appropriate scaling method for numerical features (e.g., Min-Max scaling or standardization) to ensure all features contribute equally during model training.

Encoding Categorical Variables: Choose the encoding method (e.g., one-hot encoding, label encoding) for converting categorical variables into a numerical format suitable for ML algorithms.

2. Feature Engineering Settings: Feature Selection: Determine the set of features to be included in the final dataset for model training based on their relevance and importance to churn prediction.

3. Model Pickup Options: Decide on algorithms: Identify the potential machine learning (ML) methods (such as Logistic Regression, Random Forest, Gradient Boosting, and Neural Networks) that will be tested for churn prediction.

In order to determine the best model configuration, define the hyperparameter search space and the hyperparameter tuning technique (such as Grid Search, Random Search, and Bayesian Optimization).

4. Typical Training Environments: Test-Train Split: To evaluate the effectiveness of the model's generalization, determine the ratio for dividing the dataset into training and testing groups.

Cross-Validation: When tuning hyperparameters, choose the number of folds and the evaluation metric for cross-validation.

5. Evaluation Metrics and Model Evaluation Settings Describe the key evaluation measures that will be used to evaluate the model's performance on the test set, such as accuracy, precision, recall, F1-score, and ROC-AUC.

6. Data Update and Maintenance Settings:

Data Refresh Frequency: Specify the frequency for updating the model with new customer data to ensure the model's predictions remain accurate over time.

Retraining Policy: Define the conditions under which the model should be retrained, such as a significant change in data distribution or a predefined time interval.

7. Monitoring and Logging Settings:

Logging Mechanism: Set up a logging system to record prediction requests, model performance, and potential errors for monitoring and auditing purposes.

Monitoring Frequency: Decide the frequency of monitoring the deployed model to detect anomalies or issues promptly.

8. Security and Privacy Settings:

Data Encryption: If dealing with sensitive customer data, implement data encryption during data transmission and storage.

Access Controls: Implement user access controls to impose limitations on who can access the ML system and its components in accordance with user roles and permissions.

**Data Design**

**List of Key Schemas/Tables in database:**

Establish user access controls to limit access to the ML system and its components in accordance with user roles and permissions. Customers: Each Telco customer's ID, name, gender, age, contact information, and demographic data would probably be stored in this table.

* Services: Information on the various services provided by the telecommunications provider, such as internet plans, phone plans, TV subscriptions, etc., might be found in this table. Most likely, each service would have its own service ID.
* Subscriptions: The subscriptions that each customer has chosen to join are maybe represented in this table. It would be linked to the Customers and Services tables and would include details about the services each customer is utilizing, as well as information on subscription start and end dates, billing cycles, and other things**.**
* Payments/Billing: For each client, this table would contain billing information, payment history, and information about the various payment options. Dates, sums, payment statuses, and client payment methods could all be included.
* Churn History: This table could be created to keep track of customer churn. It would include information on when a client churned (i.e., ended their subscription)

**Details of Access Levels on Key Tables in Scope:**

The privileges and permissions given to certain users or user groups for accessing particular database tables are defined by access levels. Access to the database is normally available to all Kaggle users for reading and analysis in this public dataset environment. Access levels would be necessary in a real-world situation to guarantee data security and privacy, though. Here are some examples of access levels:

* Read-Only: Users who have read-only access can only access the tables' data; they are unable to edit or delete any records. Typically, parties that require access to the data for reporting and analysis are given this degree of access.
* Read/Write: Users that have read/write access to tables can add, update, and delete records in addition to viewing data. Employees with this level of access are typically in charge of data entry or data maintenance.

**Key Design Considerations in Data Design:**

There are several important factors that must be taken into account while building the database structure and tables:

* Data normalization involves breaking up data into several related tables to cut down on duplication and boost data integrity. Data abnormalities and inconsistencies are prevented via normalization.
* Data Types and Constraints: To maintain data correctness and consistency, choose the proper data types for each column and apply constraints (such as primary keys, foreign keys, and unique constraints).
* Indexing: To enhance query performance, create indexes on columns used often for data retrieval. However, as it can affect insert and update processes, excessive indexing should be avoided.
* Performance optimization: To ensure quick response times and effective resource use, optimize database operations and queries. Data extraction and analysis for reporting and business intelligence purposes should be made simple by designing the database structure in this way.

These design factors can make the database more reliable, secure, and successful at meeting the Telco's data needs.

**Details of other frameworks being used:**

In addition to traditional machine learning frameworks like scikit-learn (Python) and Weka (Java), several other frameworks and libraries are commonly used in customer churn prediction in the telecom industry:

1. **TensorFlow / Keras:**
   * These frameworks are commonly used for implementing deep learning models, such as neural networks, which can handle complex patterns in customer behavior and churn prediction.
2. **PyTorch:**
   * PyTorch is an open-source deep learning framework developed by Facebook.
   * Like TensorFlow, PyTorch allows the creation of advanced neural network models, offering flexibility and ease of use for researchers and data scientists working on churn prediction tasks.
3. **XGBoost / LightGBM:**
   * XGBoost (Extreme Gradient Boosting), LightGBM is a popular gradient boosting frameworks that excel in handling tabular data and large-scale datasets.

**References:**

1. Iris Figalist, Christoph Elsner, Jan Bosch, Helena Holmstrom Olsson. “Customer Churn Prediction in B2B Contexts” [2019]
2. Pallav Routh, Arkajyoti Roy, Jeff Meyer. “Estimating customer churn under competing risks”. August 2020 Journal of the Operational Research Society.

1. Ahmad Hammoudeh, Malak Fraihat, Mahmoud Almomani. Selective Ensemble Model for Telecom Churn Prediction.[2019]
2. Debjyoti Das Adhikary, Deepak Gupta. “Applying over 100 classifiers for churn prediction in telecom companies.” Multimedia Tools and Applications[2020]
3. Ullah, B. Raza, A. K. Malik, M. Imran, S. U. Islam and S. W. Kim, "A Churn Prediction Model Using Random Forest: Analysis of Machine Learning Techniques for Churn Prediction and Factor Identification in Telecom Sector," in IEEE Access, vol. 7, pp. 60134-60149, 2019.
4. Ahmet Tuğrul Bayrak, Asmin Alev Aktaş, Orkun Susuz, Okan Tunalı. ”Churn Prediction with Sequential Data Using Long Short Term Memory”[2021]