**Churn Prediction and Prevention System for Subscription Services**

**Step 1: Prototype Selection**

# Problem Statement

In the competitive landscape of subscription-based services, reducing customer churn (cancellation of subscriptions) is crucial for business sustainability and growth. The project aims to develop a sophisticated Churn Prediction and Prevention System that utilizes data analytics and machine learning techniques to forecast potential churn and suggest proactive strategies to retain subscribers.

# Market/Customer/Business Need Assessment

In comprehensive evaluation of the market conditions, customer requirements, and business needs that justify the development and implementation of a churn prediction and prevention system using machine learning techniques. This assessment aims to understand the reasons for churn in subscription services and how addressing these issues can provide value to the business and its customers.

**Market Need Assessment:**

What is the overall churn rate in the subscription services industry? • Are there any emerging trends or challenges in customer retention within this market? • Are there existing solutions or competitors addressing churn in similar ways? • What potential benefits can a churn prediction and prevention system offer to the market?

**Customer Need Assessment:**

What are the primary reasons customers cancel their subscriptions? • What factors contribute to customer satisfaction and loyalty? • How can a churn prediction and prevention system enhance the customer experience? • What features or approaches would customers value in such a system?

# Target Specifications

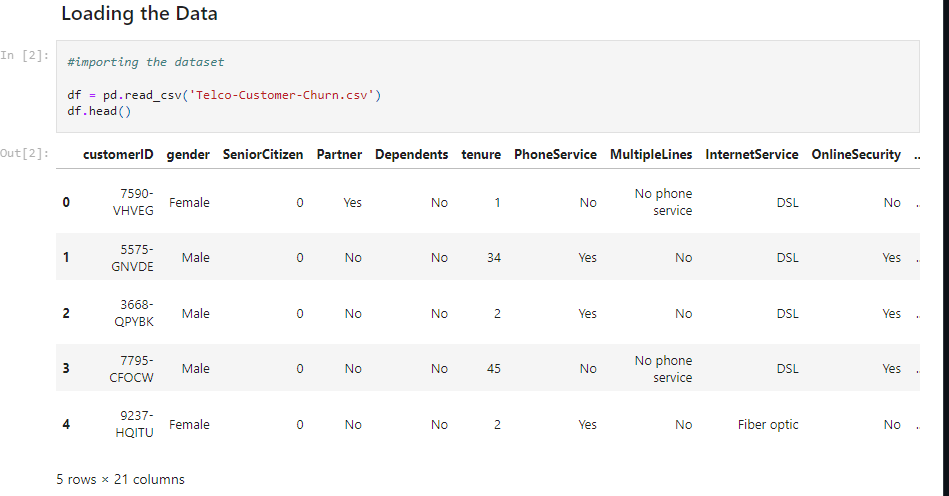
* Churn Prediction Model: The development of accurate machine learning models that can predict which customers are likely to churn (cancel their subscription) in the near future.
* Churn Prevention Strategies: Identification and implementation of strategies to prevent or mitigate customer churn based on the predictions made by the model. This could involve personalized offers, discounts, or targeted communications to retain customers
* . Model Performance Metrics: Setting clear criteria for evaluating the performance of the churn prediction model. This might include metrics such as accuracy, precision, recall, F1-score, or area under the ROC curve.
* Integration with Existing Systems: If the subscription service already has customer data and systems in place, the target specifications might involve integrating the developed churn prediction and prevention system seamlessly into the existing infrastructure.

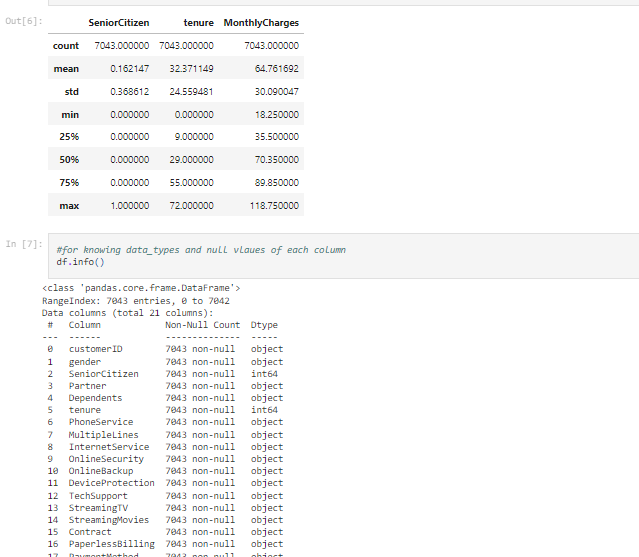
# External Search

* [dataset](https://github.com/AngadiAbhinay01/Feynn_Labs_Internship/blob/main/Project_3/Telco-Customer-Churn.csv)
* [Machine\_learning\_Algorithm](https://github.com/AngadiAbhinay01/Feynn_Labs_Internship/blob/main/Project_3/my_contribution_final_project.ipynb)

***Let’s import the dataset and have a look at it!***







# Applicable Regulations

The patents mentioned above might claim the technology used if the algorithms are not developed and optimised individually and for our requirements. Using a pre-existing model is off the table if it incurs a patent claim.

1. Must provide access to the 3rd party websites to audit and monitor the authenticity and behavior of the service.

2. Enabling open-source, academic and research community to audit the Algorithms and research on the efficacy of the product.

3. Laws controlling data collection : Some websites might have a policy against collecting customer data in form of reviews and ratings.

4. Must be responsible with the scraped data : It is quintessential to protect the privacy and intention with which the data was extracted

# Applicable Constraints

1.Continuous data collection and maintenance

2. Lack of technical knowledge for the user

3. Taking care of rarely bought products

# Business Opportunity

* Subscription Licensing
* Pay-Per-Use
* Tiered Plans
* Consulting and Integration
* Partnerships with Subscription Services
* Data Monetization (with User Consent):

# Concept Generation

o Project Understanding and Scope Definition:

o Data Collection and Preprocessing: o Feature Engineering:

o Model Selection and Architecture:

o Model Training and Evaluation:

o Churn Prediction:

o Monitoring and Continuous Improvement:

o User Interface and Visualization:

o Deployment and Integration:

This product requires the tool of machine learning models to be written from scratch in order to suit our needs. . Tweaking these models for our use is less daunting than coding it up from scratch. A well trained model can either be repurposed or built. But building a model with the resources and data we have is dilatory but possible. The customer might want to spend the least amount of time giving input data. . This accuracy will take a little effort to nail, because it’s imprudent to rely purely on Classic Machine Learning algorithm.

**Algorithms:**

# Support Vector Machine

A Support Vector Machine (SVM) is a supervised machine learning algorithm that is used for classification and regression tasks. It is particularly effective for binary classification problems but can be extended to handle multi-class classification as well. The main idea behind SVM is to find a hyperplane that best separates data points belonging to different classes.



SVMs have some key advantages:

They are effective in high-dimensional spaces.

They work well for both linear and non-linear classification tasks.

They are robust against overfitting, especially when using a proper choice of regularization parameter.

However, SVMs also have limitations, such as sensitivity to the choice of hyperparameters and the need for efficient optimization algorithms, which can be computationally expensive for large datasets.

**Tools:**

* [Python:](https://www.python.org/) It’s a programming language that will be used for building the service.
* [Pandas:](https://pandas.pydata.org/) Pandas is a library mainly used for handling, manipulating and transforming data.
* [Scikit-learn:](https://scikit-learn.org/stable/index.html) It is the gold standard library for machine learning which comes with plenty of algorithms to perform different tasks such as regression, classification etc.
* [Matplotlib](https://matplotlib.org/) and [Seaborn:](https://seaborn.pydata.org/) Both of these libraries are used for visualization purposes.