

# **CASE STUDY ASSIGNMENT**

## **Case Study -1: Retail Sales Forecasting**

### **Business Understanding:**

- **Question -1:** Key Objectives
  - To optimize inventory management across 200 + stores.
  - To meet customer needs with product availability.
  - To decrease storage expenses by maintaining sufficient amount of stocks.
- **Question -2:** Success Criteria
  - Maintain Accurate Forecasts of Sales with minimal error. Error is known as Mean Absolute Percentage Error (MAPE) in which less than 20% is considered good.
  - Decrease in overstocking which handles maintenance of holding costs and stocks.

### **Data Understanding:**

- **Question -1:** Types of data
  - Monthly Sales Data for each category of product across all stores from the beginning.
  - Each stores information with geographical and economic conditions.
  - Each product information with specifications and pricing.
  - Competitors, Market trends are also required to build model.
- **Question-2:** Quality of data
  - Check whether the data is available from all the stores and missing values need to be replaced by physical inventory. Validation is necessary.
  - Need to create a standard format of data from every store with regular updates of data timely.
  - Ensure that sales data is correct, which influences patterns while predicting.

## **Data Preparation:**

- **Question -1:** Outlining steps
  - Cleaning: Fill Missing Values, No Repetition of Data, Standardize data from every store, Regular updating of data.
  - Feature Engineering: Create time-based features (day, week, month, season), product popularity, and store-level characteristics.
- **Question -2:** Challenges
  - Main challenge is missing data as it is historical data from several years which leads to poor forecasting.
  - Handling seasonal patterns of products in which customers buy only in the festive seasons.

## **Case Study -2: Customer Churn Prediction**

### **Modelling:**

- **Question -1:** Modelling
  - Logistic Regression can help to predict a binary outcome, such as whether customers will stay or leave. This regression uses many predictors like purchase history, geographical data, customer interaction with type of streaming service.
  - Decision Tree is also suitable for prediction because a decision tree may divide the client base according to age and then subdivide each group according to purchasing patterns. The client segment will get increasingly detailed with each successive step, enabling the algorithm to forecast the group's chance of churn.
- **Question -2:** Handling Imbalanced Class
  - We can handle imbalanced classes in churn dataset by using Synthetic Minority Over-sampling technique (SMOTE). For example, if we have only 100 customers, of which 50 cancelled their service. Instead of SMOTE creating new samples finding similar cases generating new data points. It creates new points for two cancelled customers having similar features. This approach helps balance the dataset.

## **Evaluation:**

- **Question -1: Metrics for Evaluation Performance**

- AUC-ROC measures how good a model differentiated between classes.
- F1 Scores can also measure performance in which it is a balanced measure between precision (Prediction of customers who will churn) and recall(Original Customers who churned).

- **Question -2: Evaluation**

We can evaluate with these methods:

- Train-test Split with time component: Split the data chronologically (not randomly). Train on Older data, test on newer data
- K-fold cross-validation with stratification (to handle imbalanced classes). Time Series cross-validation for temporary data.

## **Deployment:**

There are various kinds of performance monitoring. There are as follows

- **Regular Monitoring:**

- Track Prediction accuracy Weekly/monthly
- Monitor data drift (are the customer behaviors shifting over time)
- Compare actual vs predicted churn rates.

- **Retraining Strategy:**

- Set up Automatic retraining when performance drops below threshold.
- Periodically updated model with new data and Maintain version control of different model iteration.
- The key is to continuously monitor and update the model to ensure it adjusts and give the results effective as customer behavior patterns change over time.

## References:

- *Mean Absolute percentage error (MAPE): what you need to know.* (2023, February 10). Arize AI. <https://arize.com/blog-course/mean-absolute-percentage-error-mape-what-you-need-to-know/>
- GeeksforGeeks. (2023, December 21). *What is Feature Engineering?* GeeksforGeeks. <https://www.geeksforgeeks.org/what-is-feature-engineering/>
- *Churn prediction explained* / Stripe. (2024, March 4). <https://stripe.com/resources/more/churn-prediction-101-how-to-choose-the-best-prediction-model-for-your-business#:~:text=Logistic%20regression%20is%20a%20statistical,whether%20a%20customer%20will%20churn.>
- Maklin, C. (2022, May 21). Synthetic Minority Over-sampling TEChnique (SMOTE) - Cory Maklin - Medium. *Medium*. <https://medium.com/@corymaklin/synthetic-minority-over-sampling-technique-smote-7d419696b88c>