NAAN MUDHALVAN

PHASE 2 PROJECT SUBMISSION

**PRODUCT SALES ANALYSIS**

**TEAM MEMBERS:**

PARVATHY A - 2021504533

PAVITHRAN E - 2021504535

SHAMEEM AHAMED S - 2021504042

THUSHIYANTH K - 2021504555

**PROBLEM DEFINITION:**

This project involves using IBM Cognos to analyse sales data and extract insights about top-selling products, peak sales periods, and customer preferences. The objective is to help businesses improve inventory management and marketing strategies by understanding sales trends and customer behaviour. This project includes defining analysis objectives, collecting sales data, designing relevant visualizations in IBM Cognos, and deriving actionable insights.

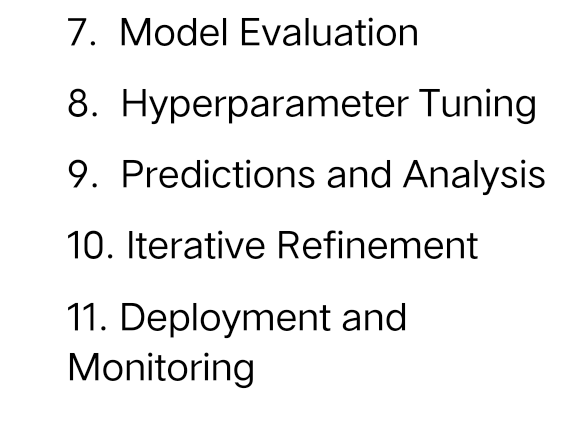
**DATABASE LINK:**

[**https://www.kaggle.com/datasets/ksabishek/product-sales-data**](https://www.kaggle.com/datasets/ksabishek/product-sales-data)

**OBJECTIVES:**

* Analysing data related to product sales.
* Generating valuable insights from the data.
* Based on the insights, recommendations must be formulated to address issues and optimize sales and profitability.

**DESIGN APPROACH:**

1. Data Collection and Preparation
2. Feature Selection
3. Data Splitting
4. Feature Engineering
5. Model Selection
6. Training the Model

**1) DATA COLLECTION AND PREPARATION:**

* The dataset we are going to work on for this project is taken from **Kaggle** and we have mentioned it above.
* We also gather other historical sales data, including information on **products, customers, prices, promotions, and time periods of a company** and work on it.

**2) FEATURE SELECTION:**

We search for relevant features in which we are going to work on that could influence sales. (**Product attributes, pricing, customer demographics, seasonality, and marketing activities**).

**3) DATA SPLITTING:**

We will split the data into training and testing sets. The training set is used to train the machine learning models, while the testing set is used to evaluate the model's performance.

**4) FEATURE ENGINEERING:**

We will create new features that might provide valuable insights. (**Date & Time** features, **Customer behaviour** features, **Promotion and marketing** features etc…)

**5) MODEL SELECTION:**

Selecting appropriate machine learning algorithm depending on the nature of the problem. In this we are going to use either one of the following algorithms,

1) Regression Algorithm (**Linear Regression, Decision Trees, Random Forest**).

2) Gradient Boosting.

3) Neural Networks.

**6) TRAINING THE MODEL:**

Then we will train the selected machine learning models

using the training dataset that we have classified before. The model learns

**patterns and relationships** between the features and the target variable (sales).

**7) MODEL EVALUATION:**

The model’s performance is evaluated using the testing

dataset. The common metrices for Regression based model training includes,

* Mean Absolute Error (**MAE**)
* Mean Squared Error (**MSE**)
* Root Mean Squared Error (**RMSE**)

**8) HYPERPARAMETER TUNING:**

Optimize the model's **hyperparameters** to improve its performance. Techniques like **Grid Search or Random Search** can be employed to find the best set of hyperparameters.

**9) PREDICTION AND ANALYSIS:**

We use the trained model to make **predictions on new data** that is not known to the model. Analyse the predictions to identify trends, patterns, and areas for further improvement.

**10) ITERATIVE REFINEMENT:**

Iterate through the process, **refining features, experimenting with different algorithms**, and tuning hyperparameters to improve the model's accuracy and reliability.

**11) DEPLOYMENT AND MONITORING:**

Once we achieved a model with **great accuracy**, we will deploy it into a sales system for making real-time applications. We will update the model by monitoring continuously its performance over time.

**FLOWCHART:**



