Business Strategic planning (Data Science Model)

In this paper, I'll be discussing the problem statement "Can we accurately predict the stock levels of products based on sales data and sensor data on an hourly basis in order to more intelligently procure products from our suppliers?" required for the company utilizing sales data & sensor data.
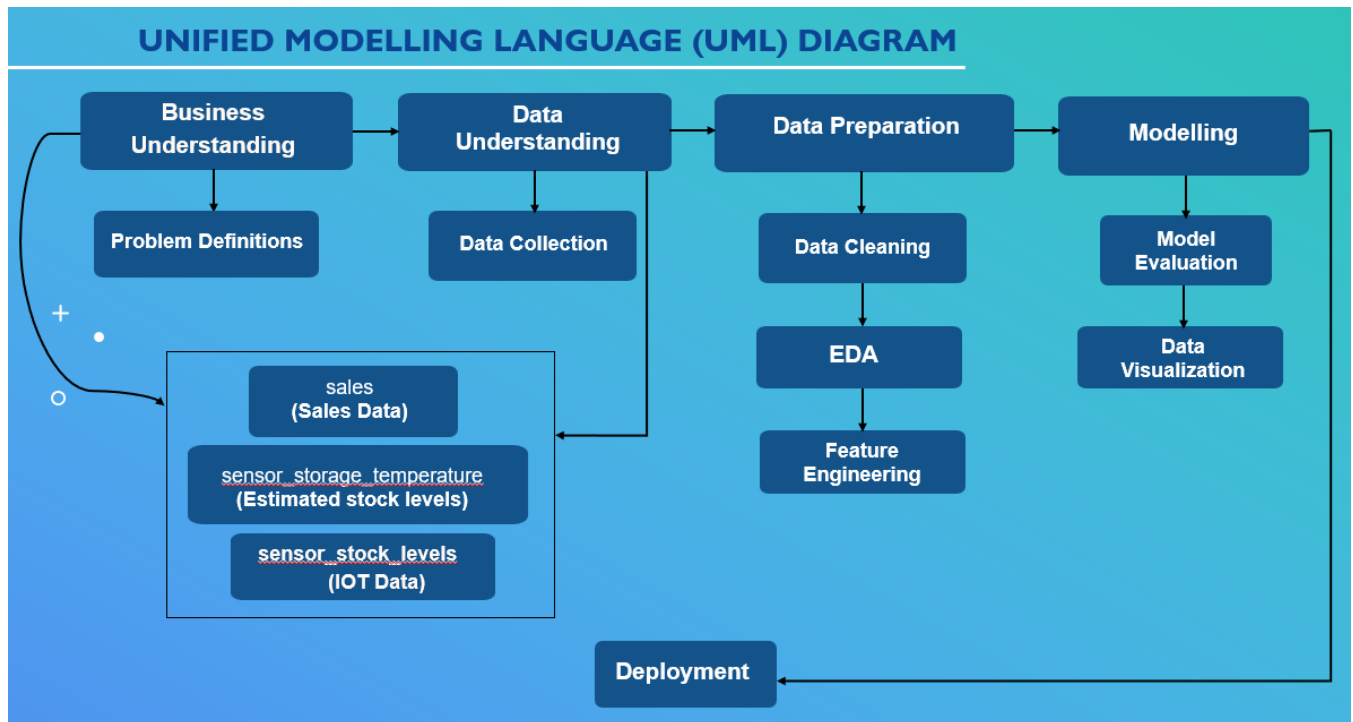

Fig-1: UML diagram of the business model

Step-1: Business Understanding (Problem Definitions)

Our aim is to solve the problem statement "Can we accurately predict the stock levels of products based on sales data and sensor data on an hourly basis in order to more intelligently procure products from our suppliers?" utilizing the company's sales & IOT sensory data.

Step-2: Data Understanding (Data Collection)

Using the data model provided to us we have build a UML diagram for the business model describing how we will proceed to solve the problem statement that the client has positioned. The client has agreed to share more data in the form of sensor data. They use sensors to measure temperature storage facilities where products are stored in the warehouse, and they also use stock levels within the refrigerators and freezers in store.
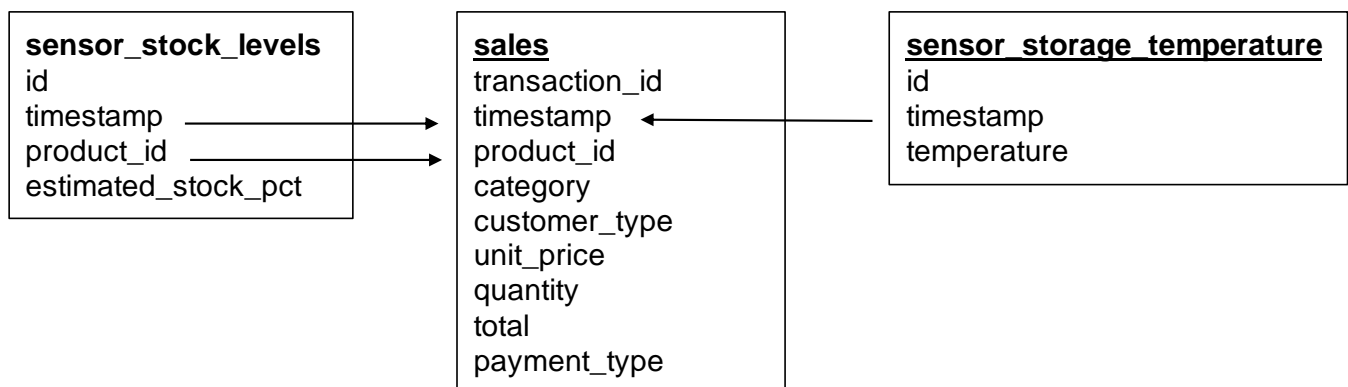
This data model diagram shows:

- sales = sales data
- sensor_storage_temperature = IoT data from the temperature sensors in the storage facility for the products
- sensor_stock_levels = estimated stock levels of products based on IoT sensors

Step-3: Data Preparation

First, we will check the information of all the dataset by printing out the data size (dimension) using 'shape()', number of records in the dataset using 'len()', number attributes in the dataset using 'len(x.columns)', dataset information using 'info()' & dataset description using 'describe()'. Afterwards, we will merge the datasets & convert the timestamp to 'hourly', and check whether there's any missing or null value in the dataframe using 'isnull()', then we will print out the numerical & categorical columns in the dataset. We will also implement feature engineering for which we will use 'pearson correlation coefficient' to check the correlation of features between each other including the target variable.

Step-4: Modelling

After data preparation, we will divide our dataset into training & testing in which we will train our model using specific features & will select a target variable based on problem statement. We'll make some prediction models like Decision Tree, Random Forest or Boosting models for accurate prediction of the target variable and will also calculate the F1-score, Accuracy, TP-Rate (Recall), PR-AUC (Precision-recall curve) & ROC-AUC (Receiver Operating Characteristic Curve) including error (ex- MAE (Mean Absolute Error)) in our model. Furthermore, we will visualize several models based on their performance matrix and choose the best one.

Step-5: Deployment

After making a prediction model, we can deploy our model for business purposes on various platforms as a mobile application or web application.

To conclude, we will use the UML diagram and proceed step-by-step for predicting our business problem statement for better productivity.