**PHASE 2**

**INNOVATION**

|  |  |
| --- | --- |
| **DATE** | 10 OCTOBER 2023 |
| **TEAM ID** | 715 |
| **DOMAIN NAME** | DATA ANALYTICS |
| **PROJECT NAME** | PRODUCT SALES ANALYSIS |
| **MAXIMUM MARKS** |  |

**INTRODUCTION:**

The main objective of this phase is to recognize the machine learning algorithms used for predicting the future sales and customer preferences in product sales analysis.

**PROBLEM STATEMENT:**

Predict future sales performance for each product, region depending on top-selling products, peak sales analysis and customer preferences.

**PRODECURE:**

The steps that will be followed for prediction of future sales and customer preferences by our team will be:

1. Load the dataset (available in skillup portal)
2. Prepare the data
3. Choose machine learning algorithm for prediction
4. Train the model
5. Test the model
6. Make predictions

Prepare the data

Import/Load the dataset into Jupyter Notebook

Getting sales dataset from Kaggle(Provided by IBM)

Separation of training and test data

Identifying ML algorithm-Classification algorithm

Predict the sales of product using algorithms

Test the model using remaining data.

Train the model using data.

Visualization of Data using Tools.

**LOAD THE DATASET:**

The dataset that is provided by IBM will be downloaded from Kaggle and the dataset will be imported. Jupyter Notebook will be used by our team for prediction of future sales and customer preferences. Python libraries like Numpy and Pandas will be used here.

**PREPARE THE DATA:**

The data which is imported in Jupyter Notebook will be cleaned and transformed for better accuracy. The raw data that is imported will be shaped according to the necessity. Similarly an analysis need to be done in order to define whether the data comes under classification or regression analysis.

**CHOOSE THE MACHINE LEARNING ALGORITHM:**

According to the dataset provided by IBM, the data are categorized. The dataset of sales data contain information like order priority which is categorized into high, low, critical and not specified. Therefore classification algorithms can be used for predicting future sales and customer preferences.

* **Logistic regression:** Logistic regression is a simple but powerful algorithm that can be used to model the probability of an event occurring. It is often used to predict whether a customer is likely to make a purchase, or not, whether or not a product is likely to sell.
* **Decision trees:** Decision trees can also be used for classification tasks. They work by constructing a tree-like model that represents the decisions and relationships between variables. Decision trees can be used to predict sales and customer preferences by considering multiple factors that impact these outcomes, such as customer demographics, product popularity, and past purchase history.
* **Support vector machines (SVMs):** SVMs are a type of classification algorithm that can be used to find a hyperplane that separates the data into different classes. SVMs are often used for classification tasks where the data is high-dimensional and sparse.
* **Random forests:**Random forests are an ensemble learning algorithm that combines multiple decision trees to produce a more accurate prediction. They are often used for classification tasks, such as predicting customer preferences or product demand.

**TRAIN THE MODEL:**

A model will be trained according to the data using python libraries like sklearn. Our team will train the model according to classification algorithms. After feeding the model that contain input and output data, the model learns to predict the output data for new input data.

**TEST THE MODEL AND MAKE PREDICTIONS:**

The data which is not used for training a model will be used for testing. Prediction of data will be done by using the metrices like accuracy, precision etc. When the model is performing well ,this model will be deployed and ensure that the prediction model is accurate and reliable.

**CONCLUSION:**

Overall, our team will focus on predicting future sales and customer preferences using classification algorithms like Decision Trees, Logistic Regression and produce a prediction model with greater accuracy which can help businesses to plan and make better decisions that lead to greater profit.