|  |  |  |
| --- | --- | --- |
| **Project Overview** | **Project Name: Detecting the Product and sales with Electric plug Dataset in the Marketplace Product** | **Project Manager :**  **Shefali Kamalnakhawa** |

**STATEMENT (POS)**

**Problem/Opportunity :**

In this project I will be working with Marketplace product data from **ASIN.**  The dataset covers the product as plug or no plug and predicts the sales of the product.

**Goal:**

The goal of this project is to predict the **target\_label** field (plug or no plug) of the electric plug dataset and to predict the **target\_sales** field to know the most sales of a product in marketplace using machine algorithm. From the large dataset we will find which product is plug or no plug. After detecting the product, we will predict which product is selling the most and calculate monthly revenue.

**A) Dataset:**

Dataset Link: <https://drive.google.com/file/d/1gQUSI0acSyPS5owEGO1xEZF3YMFZlgPH/view?usp=sharing>

**Dataset schema:**

* **ASIN**: Product ASIN
* **target\_label:** Binary field with values in {0,1}. A value of 1 show ASIN has a plug, otherwise 0 (the predict field).
* **ASIN\_STATIC\_ITEM\_NAME:** Title of the ASIN.
* **ASIN\_STATIC\_PRODUCT\_DESCRIPTION:** Description of the ASIN
* **ASIN\_STATIC\_GL\_PRODUCT\_GROUP\_TYPE:** GL information for the ASIN.
* **ASIN\_STATIC\_ITEM\_PACKAGE\_WEIGHT:** Weight of the ASIN.
* **ASIN\_STATIC\_LIST\_PRICE:** Price information for the ASIN.
* **ASIN\_STATIC\_BATTERIES\_INCLUDED:** Information whether batteries are included along with the product.
* **ASIN\_STATIC\_BATTERIES\_REQUIRED:** Information whether batteries are required for using the product.
* **ASIN\_STATIC\_ITEM\_CLASSIFICATION:** Item classification of whether it is a standalone or bundle parent item etc.
* **ASIN\_STATIC\_DATE:** Information on which date the product get sold.
* **ASIN\_STATIC\_ITEM:** Information about no. of items.
* **target\_sales:** Information about no. of sales of product (the predict field).

**B) Methods, Models and Algorithms:**

**I. Exploratory Data Analysis:**

Performing simple statistics on dataset to visualize the data and get to know the data in detail. Visualizing the features of the dataset. Build the model and classifier including all the features in the datasets.

**II. Data Preprocessing (Cleaning the dataset):**

Before performing Analysis clean the dataset. Check if there is any missing value, if it is then drop or replace the missing values. Clean the text data before processing the dataset.

**III. Data Processing (create a pipeline):**

Build a pipeline to impute the missing values with the mean using sklearn SimpleImputer.

**IV. Train – Validate – Test:**

Train, Validate and Test the Dataset to get the results.

**C) Tools and Python libraries:**

Here’s the list of tools we’ll use for the Interpreting data:

* Google Colab for the development environment
* Google Drive to store the data

They’re free with a basic Google account and will help keep things simple.

As for Python libraries, here’s what needed

* Pandas, Scikit-learn, nltk, matplotlib
* google.colab to link Google Drive to the Colab notebook
* csv for loading and saving csv files

**Objectives:**

The purpose of the project is to analyze, visualize and identify the patterns and trends of data to solve the business problems. Loading the dataset, analyzing the data, creating pipeline, building a classifier, apply machine learning algorithm and make predictions. In this I will discover the patterns in the dataset to predict the complex patterns which will answer business questions, detect, and analyze the trends and patterns to solve the business insights.

**I. Project will address below business problems:**

1. Predict if the product is plug or no plug?
2. Predicting which product is selling the most?
3. Predicting monthly sale of a product?

**II.** **Schedule**

|  |  |  |
| --- | --- | --- |
| Date |  | Tasks to be Completed |
| 09/22/21 |  | Data Sourcing and selecting the topic for the project |
| 10/06/21 |  | Submission of project proposal |
| 10/13/21 |  | Literature Review |
| 10/18/21 |  | Creating Presentation |
| 10/26/21 |  | Working on codes to interpret data |
| 11/05/21 |  | Working on Poster |
| 12/22/21 |  | Final Presentation, Poster, and code submission |

**Success Criteria:**

The project will solve the business problem of the companies and let them predict the monthly sales and profits they get from the products. This will analyze the trends and pattern of the product which let the company know the customers requirements. Based on that they could showcase more products related to the dataset and increase their revenues. I will get hands-on-experience with machine learning algorithms that will help in predicting and solving real life business problems.

**Assumptions, Risks, Obstacles:**

The Data quality is poor and to overcome this problem I need to reduce attributes in the dataset and simplify the dataset to perform methods and solve the problem. The accuracy rate of prediction may or may not be accurate due to specific products which will risk in calculating exact profit the company is getting by the product.

|  |  |  |  |
| --- | --- | --- | --- |
| **Prepared By** | **Date** | **Approved By** | **Date** |
| Shefali Kamalnakhawa | 9/28/21 | Dr. Christelle Scharff |  |