**[Navyashree Budhihal Mutt ​​, Kunal Nagdev, Aishwarya Sachchidanand Muchandi, Venkata Manideep K , Rahul Mogillapalli]**

**Koganti, Ramakrishna Prasad**

**CSE 5301**

**12 December 2021**

**Product Quality Check.**

**Abstract:**

People are finding it difficult to obtain essential things at low prices due to rising inflation in this expensive environment. Some organizations are willing to assist such people; however, the main flaw of these organizations is the quality of the products that they are willing to sell or donate. To help organizations overcome this difficulty, we're developing a prototype that will assist them in keeping a quality check on any product they intend to put into operation. The quality, such as date of manufacture, material, durability, pre-use history, maintenance, repairability, as well as many more characteristics are determined in numerous factors. We propose to identify the chance of continuing use of used products with all of these characteristics. Ikea Furniture Dataset will be used for the training model. By implementing ML models, supervised learning should be more accurate to anticipate. Use of the Hyper-Parameter Tuning, a comparative algo analysis (SVM, Naive Bayes Algo, etc.) will help to discover the algo most appropriate for the problem statement. The method outlined above will assist in deciding whether or not the product is consumer-friendly.

**Keywords:** ML- Machine Learning, Product Quality, Supervised Learning, Hyper-Parameter Tuning, Linear Regression, Logistic Regression.

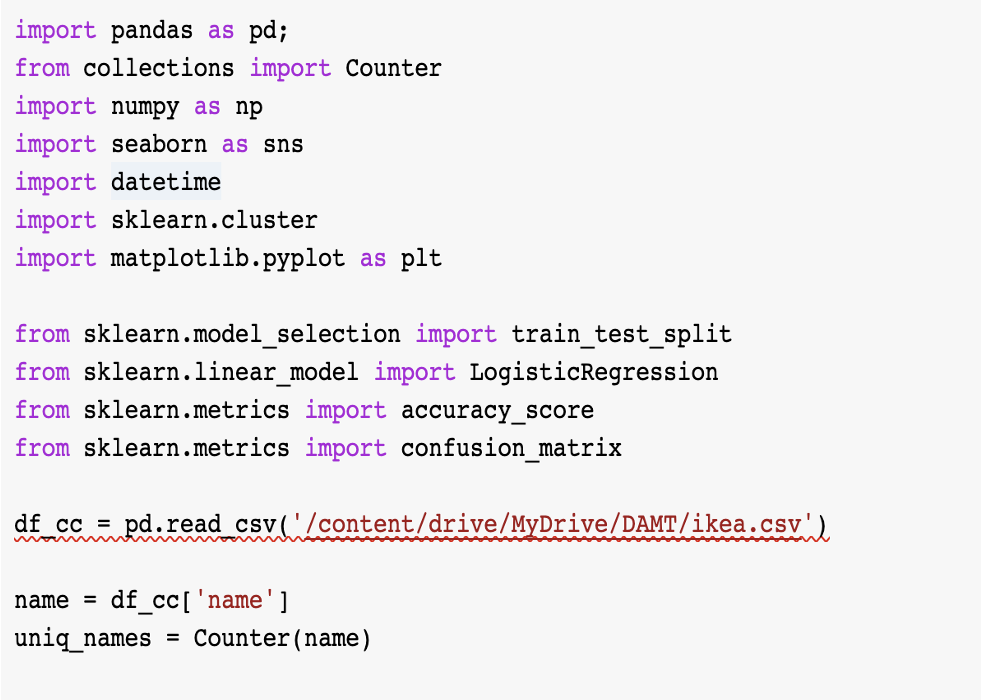
**Introduction:**

Due to product quality difficulties, a consumer is always perplexed when attempting to purchase previously used things. Our project provides a solution to such a conundrum and assists him in reaching a conclusion. The quality of the product is mostly determined by specific characteristics such as Price, other colors, short description, frame quality type, strength, flexibility. Each parameter is given a certain weight based on its significance in the creation of the product. To begin, a product analysis will be conducted to determine whether the product contains those parameters or not, as well as the condition of the quality determining parameters. Our data set can provide a complete insight of the product parameters. If the product contains such parameters, the weights for those parameters will be combined together, and the total weight will be calculated. The product's quality will be determined by its total weight. For suppose if a consumer selects a product its quality is predicted based on the parameters it possesses.

**Analysis:**

**Data collection:**

Collecting data for training the ML model is the basic step in the machine learning pipeline. The predictions made by ML systems can only be as good as the data on which they have been trained. Here the dataset used in the program is Ikea.csv where this dataset is obtained from the kaggle. The data collection helps in collecting the structured data and helps in extracting the features from data to train the machine learning model. From the given figure below the data obtained from kaggle is imported into the code. As this is the initial step for the Project.



**Fig 1 [** importing the CSV file obtained from the kaggle to train the machine learning model]

**Data cleaning and preprocessing the data:**

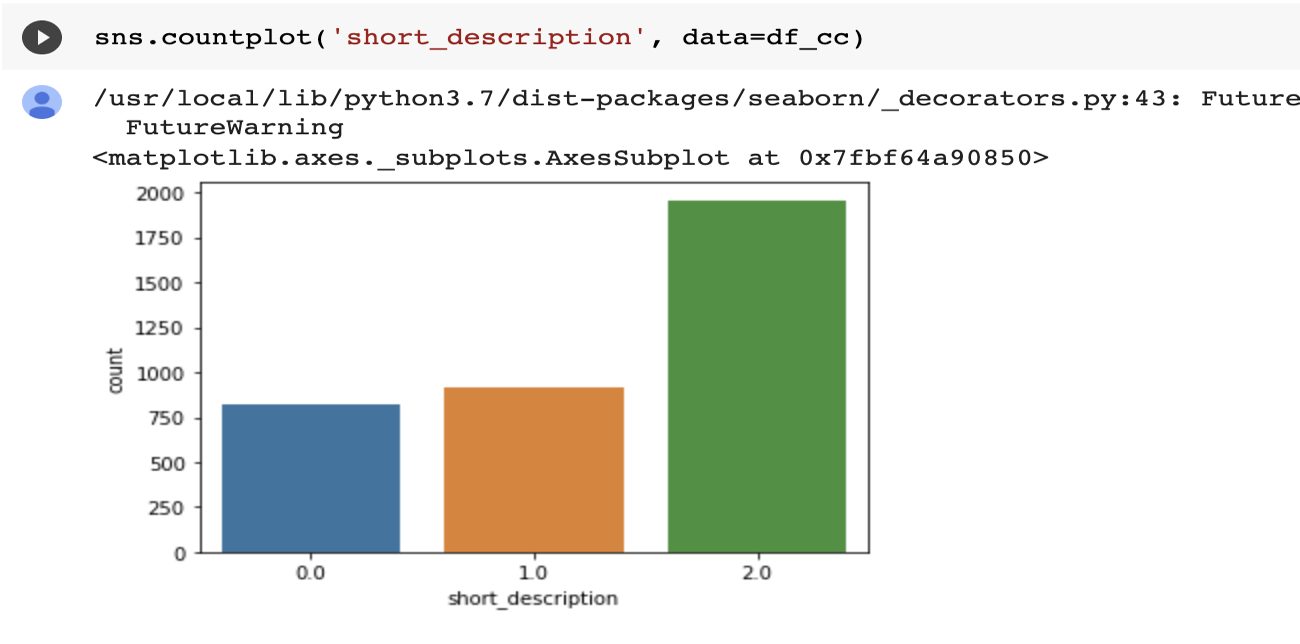
The next step after data collection is data cleaning and data preprocessing is data cleaning and data preprocessing. Performing this step helps the data become more useful and this will directly affect the model to be able to perform operations over the dataset more accurately.Real-world raw data are often incomplete, inconsistent and lacking in certain behaviors or trends. So, once collected, they are pre-processed into a format the machine learning algorithm can use for the model. Data cleaning is techniques, manual and automated, removing data incorrectly added or classified. While preprocessing the dataset the data set might contain the some missing data, incorrect data which will be a problem while training the data to get the accurate values, so while performing the preprocessing the data



**Fig 2 [ collecting the required data for training the model from the CSV file]**

**Data wrangling for clear data understanding**

The first step in data understanding is a Data Assessment, exploratory Data Analysis. This should be undertaken before the Start of a project as it is an important step to validate its feasibility. exploratory data analysis is an approach of [analyzing](https://en.wikipedia.org/wiki/Data_analysis) [data sets](https://en.wikipedia.org/wiki/Data_set) to summarize their main characteristics, often using [statistical graphics](https://en.wikipedia.org/wiki/Statistical_graphics) and other [data visualization](https://en.wikipedia.org/wiki/Data_visualization) methods. It helps us to decide which models can be used and which can not be used. This is a phase for creating meaningful summaries out of your data and is particularly important if you are unfamiliar with the data and gain basic understanding.





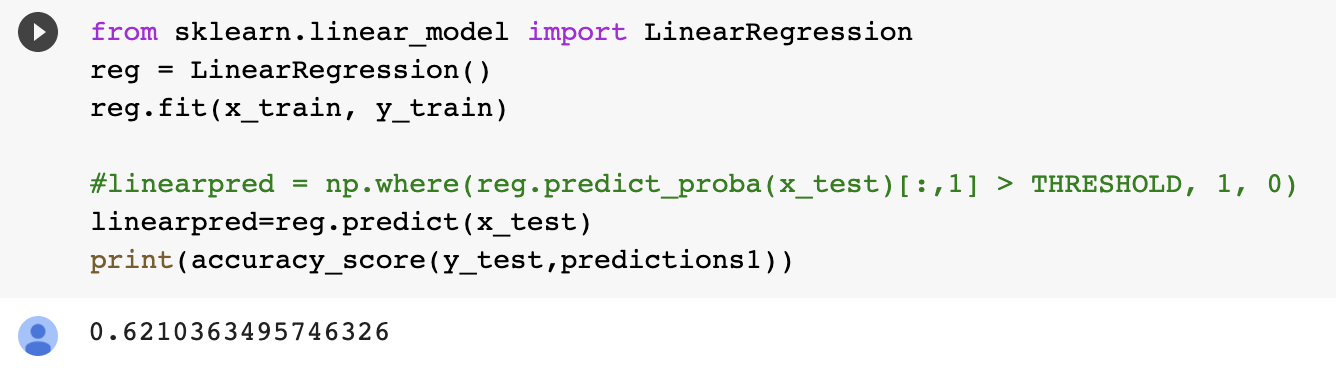
**Fig 3 [ above picture is the graphical representation of the collected and preprocessed data from CSV file.]**

**Training the model:**

Machine Learning algorithms learn from data and develop generalisation and understanding, make decisions, and we evaluate their confidence from the training data they’re given with the help of evaluation metrics such as accuracy score, etc. , better the training data is, the better the model performs. For our project we have employed Logistic Regression as our model and accuracy score as evaluation metric.

**Linear regression:**

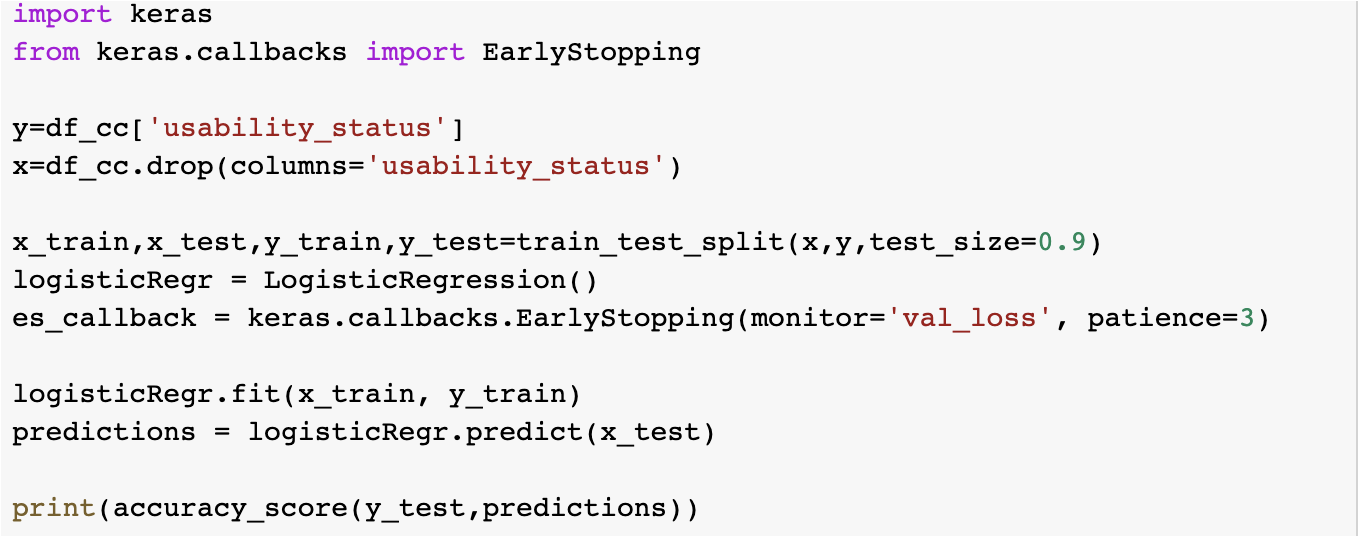
The most common algorithm used for the prediction in machine learning is linear regression, this algorithm predicts the usability of the furniture based on the variable dataset equation. It gives a statistical approach for the known problem. It helps build the relationship between the independent and dependent variables to calculate the accuracy and work on the known problem depending on the accuracy.



**Fig 4 [ above figure denotes the linear regression and the accuracy obtained]**

**Logistic Regression:**

The logistic model is a statistical model used to model the probability of a certain class or event existing such as pass/fail, win/lose, alive/dead or healthy/sick. In our case it is classes usable / not usable which is similar to a binary classification task.



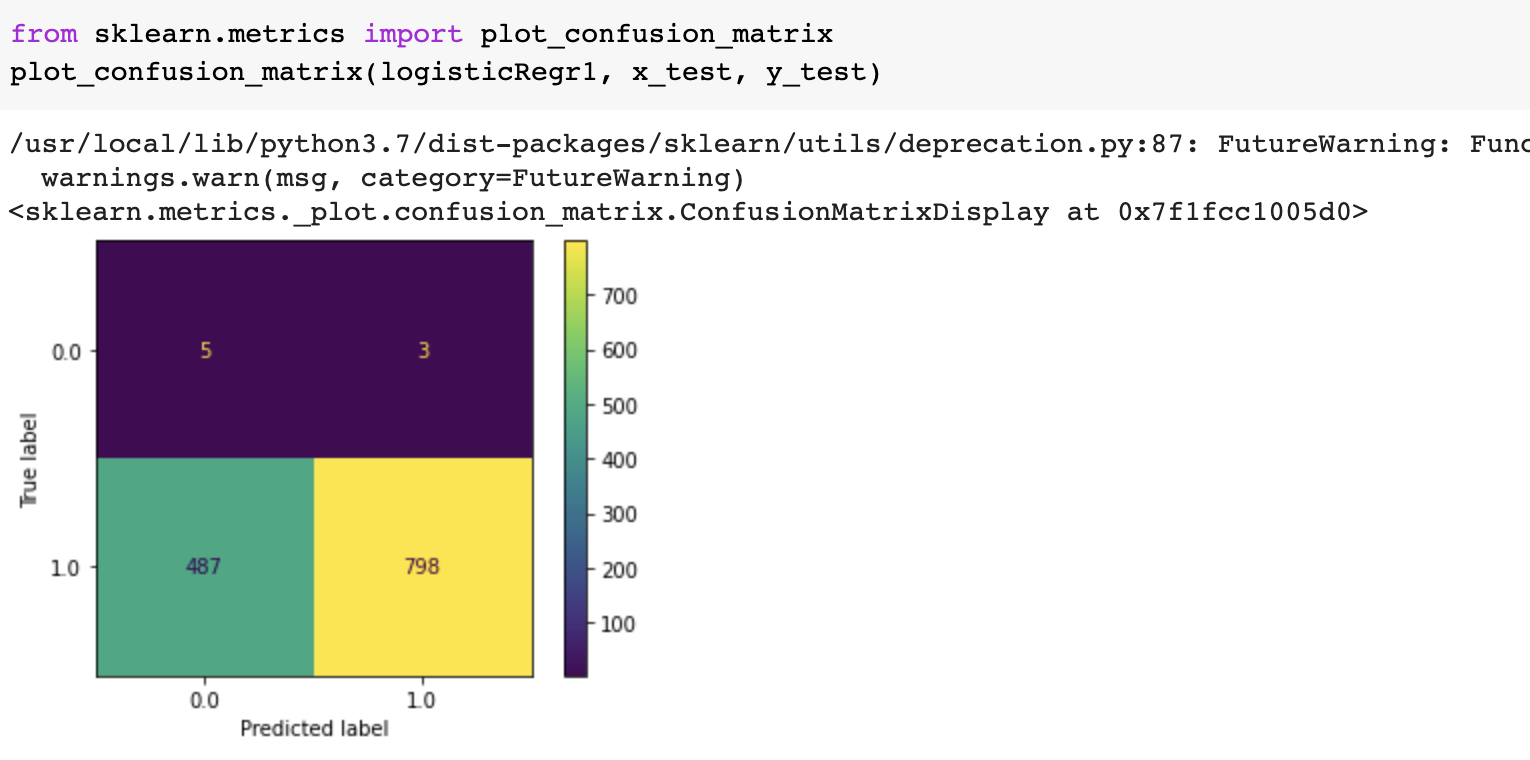
**Fig 5 [ above figure demonstrates the logistic regression training model]**

**Results:**

Following the completion of the analysis, the final result is obtained. If a consumer chooses a specific product, he will be able to determine whether the product quality is good or not and can draw a final conclusion about the product based on the output generated.



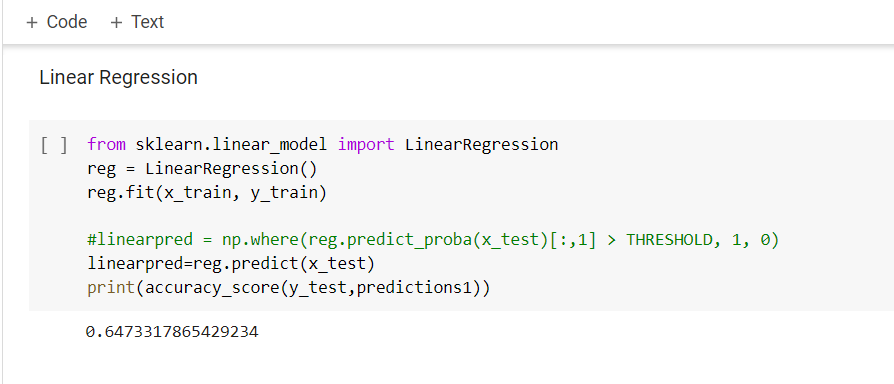
**Fig 5 [ ikea CSV dataset which is used to train the model]**

****

**Fig 6 [ confusion matrix obtained from the logistic regression]**

****

**Fig 7 [Logistic regression accuracy result]**

****

**Fig 8 [ Linear regression accuracy result]**

**Conclusion:**

After completion of data pre-processing, dataset will be split into train and test. Using trained data, planned to train, model algorithms to predict the quality of the product and also determine the price of product based on the train dataset. Finally, passing the test dataset, to predict the prices as per training the model.

**Reference:**

1. Url: https://keras.io/api/

2. Url: https://www.kaggle.com/ahmedkallam/ikea-sa-furniture-web-scraping