

MES COLLEGE OF ENGINEERING, KUTTIPPURAM  
DEPARTMENT OF COMPUTER APPLICATIONS  
20MCA246 – MAIN PROJECT

---

**PRO FORMA FOR THE APPROVAL OF THE FOURTH SEMESTER MAIN PROJECT**

---

*(Note: All entries of the pro forma for approval should be filled up with appropriate and complete information. Incomplete Pro forma of approval in any respect will be rejected.)*

Main Project Proposal No : \_\_\_\_\_  
(Filled by the Department)

Academic Year : 2021- 22

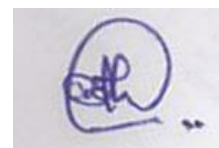
Year of Admission : 2020

1. Title of the Project : BIG MART.
2. Name of the Guide : MS. PRIYA JD
3. Student Details (in BLOCK LETTERS)

Name

Register Number

Signature



MOHAMMED ADHEEB P

MES20MCA-2025

Date:

**Approval Status :** Approved / Not Approved

Signature of  
Committee Members }

---

**Comments of the Guide**

Dated Signature

Initial Submission :

First Review :

Second Review :

---

**Comments of the Project Coordinator**

Dated Signature

Initial Submission:

First Review

Second Review

Final Comments:

Dated Signature of HOD

## **Introduction:**

E-commerce or electronic commerce is a process of Buying, selling, transferring, or exchanging products, services, and/or information via electronic networks and computers. Commerce is a fundamental to the success of any business. To streamline trading operations and maintain profits, the industry must focus on commerce, which deals with a lot more than just buying and selling. As the world is digitalizing, ecommerce solutions are increasingly becoming common. The advent of machine learning and artificial intelligence has further enhanced the effectiveness of ecommerce. The examples of ecommerce include eBay, Amazon, Upwork, Olx, etc.

The advantages of ecommerce shopping are Lower price for products compared to traditional shopping, Time saving, we can avoid crowds in shopping, Available wide range and verity of products, shipping/delivery options and also it provides Feedback option from customers.

The ecommerce provides two options for selling products they are Ecommerce Website and Ecommerce Mobile Application. Some customers choose to buy goods from the website and others from the mobile app.

The company needs to make sure which one gets the most revenue/annual income for future benefits.

The annual income of a company is the total money it earns from its business operation during one financial year. I can calculate the annual income by adding the following components:

- Income from selling products or services
- Capital gains from selling the shares of the company
- Income from selling surplus equipment or property of the company
- Income from interest or fees levied on the company's intellectual property.

## **Objectives:**

The objective of this project is which variables (website or mobile app) are contributing towards the more annual income prediction based on Ecommerce Customers csv file from the company.

Firstly I use **Exploratory Data Analysis (EDA)**, it is an approach to analyse the data using visual techniques. It is used to discover trends, patterns, or to check assumptions with the help of statistical summary and graphical representations.

The next step is Data Preparation, which includes

- Data cleaning:- cleaning of data by checking for null values/missing values (if we have null values we need to impute mean, median or mode values)
- checking for duplicates and conversion of data types
- Feature engineering :- elimination of unwanted features , feature addition(if needed) and feature transformation
- Data transformation :- I need to transform the data to get good model(with low errors)

Then I test the data by using various Testing strategies, they are Normality Testing, Multicollinearity testing and also checking the outliers and Relationship between features and target variables. Finally I realize which variables are contributing towards the more annual income prediction. And split the data set for building models from it.

Then I build 4 models (Linear Regression,, Lasso, ElasticnetCV, Ridge Regression ) from it, and find out the errors ( MAE - Mean Absolute Error, RMSE - Root Mean Square Error ) from these models And also again build these same model and find out errors after the aggregation log transformation are applied to these models for getting the accurate result.

Then I finally get the best model out of these models with low MAE & RMSE on normal models and other transformation models. And build a page for annual income prediction for a new record ( Page-Python raw file, working through Sreamlit ) So that I can predict the annual income of a new record by giving the values to the variables of the dataset.

### **Problem Definition:**

A project with an Ecommerce company based in London sells clothing online but they also have in-store style and clothing advice sessions. Customers come in to the store, have sessions/meetings with a personal stylist, then they can go home and order either on a mobile app or website for the clothes they want.

The company is trying to decide whether to focus their efforts on their mobile app experience or their website. They've asked to help them figure it out.

I worked with the Ecommerce Customers csv file from the company. It has Customer info, such as Email, Address, and their color Avatar. Then it also has numerical value columns:

- Avg. Session Length: Average session of in-store style advice sessions.
- Time on App: Average time spent on App in minutes
- Time on Website: Average time spent on Website in minutes
- Length of Membership: How many years the customer has been a member.
- Yearly Amount Spent: The total amount the customer is spending.

Business problem: Interpret which variables are contributing towards the more annual income prediction

### **Basic functionalities:**

Interpret which variables are contributing towards the more annual income prediction in an ecommerce company based on the Ecommerce Customers csv file from the company (Attributes -> Avg Session length, Time on App, Time on Website, Length of Membership, Yearly amount spent).

### **Tools / Platform, Hardware and Software Requirements:**

#### **Hardware Requirements**

Processor: i5 min

Hard Disk: 100 GB or more

RAM: 8 GB

#### **Software Requirements**

Language: Python

Front End: Python-Jupyter

Operating system : windows 8 or above

IDE : Anaconda Navigator

Dataset: 2020 sales result

Techniques Used: Linear Regression, Decision Tree, Lasso, ElasticnetCV, Ridge Regression.