IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 3, November 2022
Impact Factor: 6.252

Sales Prediction using Linear Regression

V. Iswarya Kumari[1], M. Bhavya[2], N. Kasi Mounika[3], M. Yoshitha[4]
B.Tech Students, Department of Information Technology[1,2,3,4]
Prasad V Potluri Siddhartha Institute of Technology, Vijayawada, Andhra Pradesh, India

Abstract: *This sales prediction uses linear regression to predict sales in the future based on past data available and provides GUI(Graphical User Interface) for better visualization to the user. It predicts the number of sales for the number of advertisements through media like TV, newspaper, and social media. It is trained with the previous data. Sales forecasting is particularly important in industries because of the limited shelf-life of many of the goods, which leads to a loss of income in both shortage and surplus situations. Too many orders lead to a shortage of products and still too few orders lead to a lack of opportunity.*

Keywords: Data Set, Linear Regression, Training, Testing, NumPy, pandas

I. INTRODUCTION

Knowing the demand for products is needed for any company to determine whether to decrease or increase production. The simpler approach for sales prediction is using machine learning techniques to automatically develop sales predictive models rather than manual prediction. As the data is abundant and constant changes are present, linear regression is preferable to use. The dataset is divided into testing data and training data. Training data is for training the model with past data and that is evaluated with test data. Generally, 80% of data is used for training and the remaining 20% is for testing.

II. PROPOSED SYSTEM

The methods of forecasting used for planning purchasing require materials, inventory management, production, work hours scheduling, advertising, and often more, and the traditional methods were primarily focused on experienced employee opinions or statistical analysis of past data. This project predicts sales based on advertisement as it is one of the most influential factors in sales. We are using machine learning for prediction which gives us the most accurate results and determines underlying trends. Linear regression is used as it is adaptable to data changes constantly. Linear regression uses both single-linear regression and multi-linear regression. So we can give single attribute input in single linear regression while we can give more than one for multi-linear regression. Also, there is GUI displayed for the user to understand better and with ease.

III. TECHNOLOGIES USED

3.1 Jupyter Notebook

Jupyter Notebook is the original web application for creating and sharing computer documents. It offers a simple and streamlined document-centric experience. Jupyter Notebooks are very powerful tools for interactively developing and presenting data science projects.

3.2 Python Libraries

There are several libraries used in the data visualization of this application. The number is used for array operations and matrices. Pandas library is used widely used for data analysis and ML tasks. matplot library is used for creating static, animated visualizations. Seaborn is data visualization with matplot library. Plotly supports over 40 unique chart types for statistical analysis. Pickle is used in serializing and deserializing objects.[2]

3.3 Data

The dataset [1] is taken in the form of a single excel sheet or CSV table. Then the data can be divided into testing and training data. Normally, 80% of data is used for training, and the remaining 20% is used for testing.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 3, November 2022
Impact Factor: 6.252

IV. RESULTS

	socialmedia	tv	newspaper	sales
0	230.1	37.8	69.2	2210
1	44.5	39.3	45.1	1040
2	17.2	45.9	69.3	930
3	151.5	41.3	58.5	1850
4	180.8	10.8	58.4	1290

Fig:1.1 Sample dataset that is given as input

Rows indicate the amount of money and the columns except the last one indicates what source is used for the advertisement. The last column represents the number of sales according to the amount spent on advertisements. The data in figure 1.1 is used for training and testing later on.

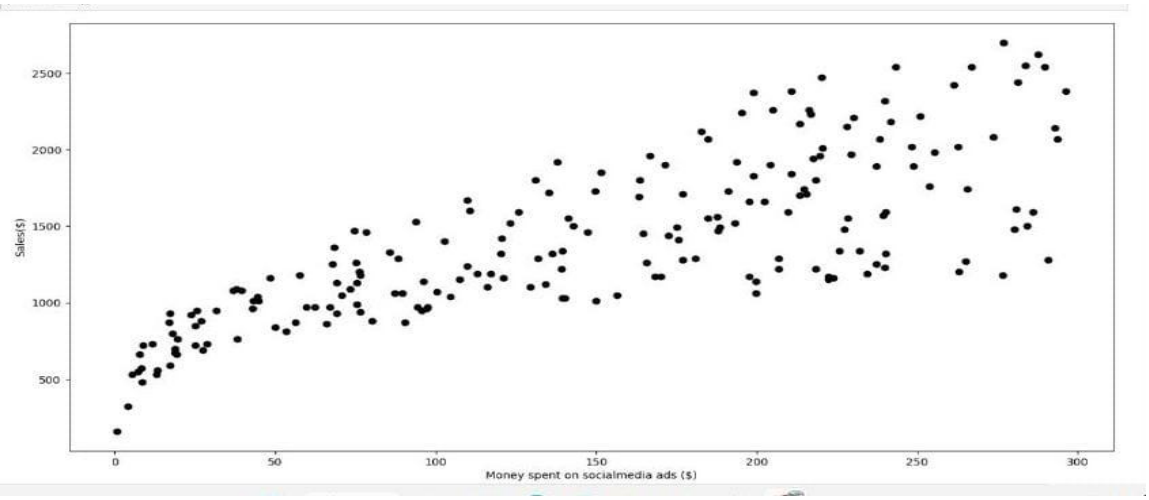

Fig:1.2 Graph for TV sales using simple linear regression to see the points

Fig 1.2 shows the real-time graph[5] for the given data set. We are using simple linear regression here. And we are only taking social media datasets. On the x-axis, money spent on social media is taken and on the y-axis number of sales is taken. We can see the points for every input of advertisement.

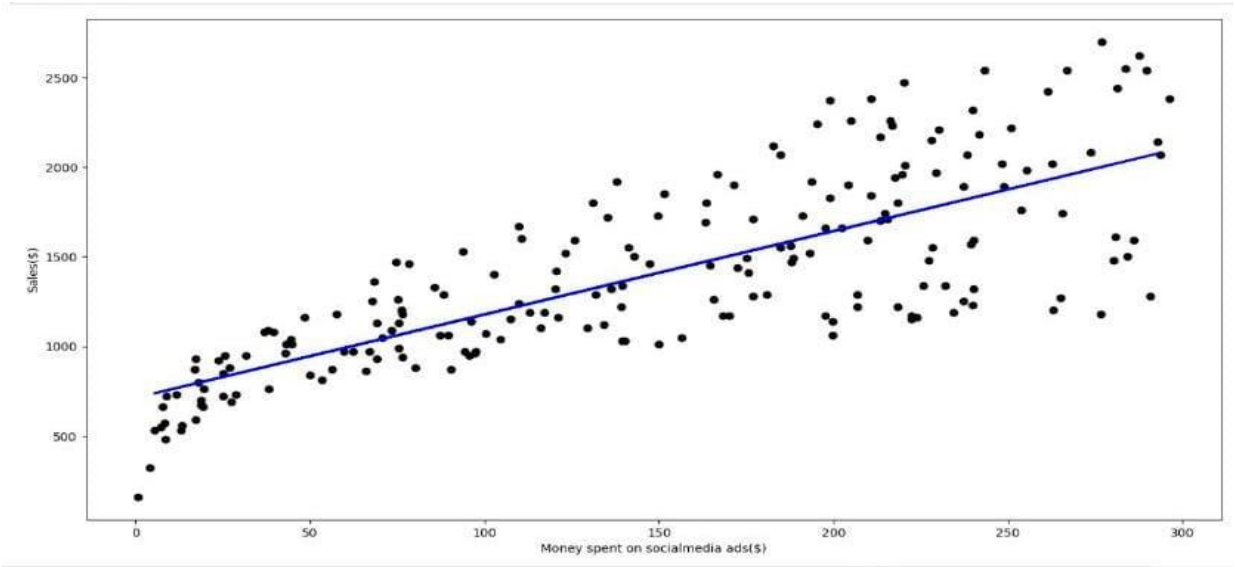

Fig:1.3 Graph for predicting the test result

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 3, November 2022
Impact Factor: 6.252

The blue line is the linear regression line. It is the best fit line for the data points. The algorithm that we are using is well known for getting the best fit line. As we are using simple regression we are giving only one attribute as input so we get a prediction for only social media advertisements.[5]

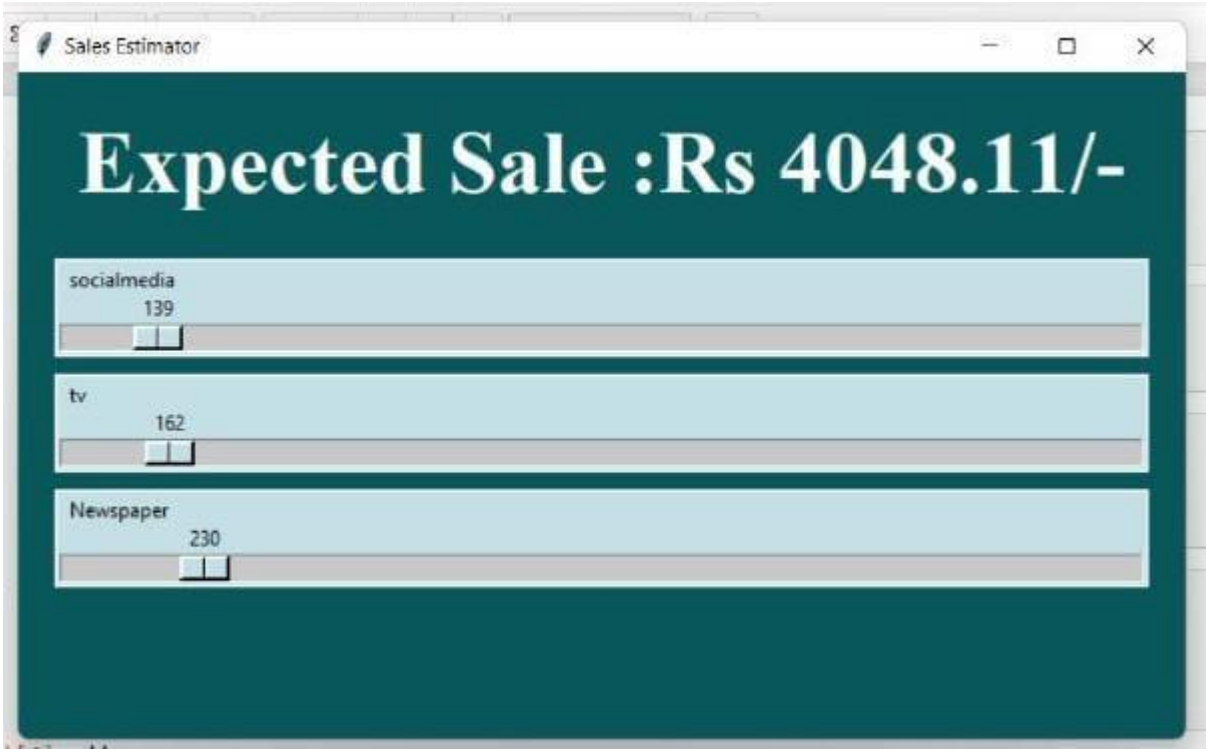

Fig:1.4 Graphical User Interface

Fig 1.4 represents the Graphical User Interface[4] which shows expected sales for the amount of money invested in advertisements. Here, we are using multiple linear regression so for the amount of money we are investing in TV, social media, and newspaper we get the expected sales. The user need not type the number but can just drag the bar and the output will be displayed according to the input.

V. SCOPE OF FUTURE USE

The application can be developed further by providing more accurate results along with ease of usage. Stock prediction can also be done if this is developed accordingly. Also, the predictions can be made not just for advertisements but also based on other factors that influence sales.

VI. CONCLUSION

Therefore, we can predict sales based on the amount spent on advertisements. As there might be huge data for input we are using a logical regression algorithm and it is a better algorithm for adapting to changes in the dataset. Also, we displayed GUI for the user's convenience.

REFERENCES

[1]. Datasets–https://www.kaggle.com/code/ashydv/sales-prediction-simple-linear-regression/data
[2]. https://www.geeksforgeeks.org/python-linear-regression-using-sklearn
[3]. Notebook- https://www.kaggle.com/docs/notebooks
[4]. https://www.geeksforgeeks.org/python-gui-tkinter
[5]. https://towardsdatascience.com/graphs-and-linear-regression
[6]. https://jupyter.org/try-jupyter/retro/notebooks/?p zath=notebooks/Intro.ipynb