SALES PREDICTION BASED ON THE AMOUNT SPENT ON ADVERTISEMENT

Sugandan K. M[1], Jeevarathinam A[2]

[1]Student, Department Of M.Sc Computer Science, Sri Krishna Arts and Science College, Coimbatore, India.

[2]Assistant Professor, Department Of Computer Science, Sri Krishna Arts and Science College, Coimbatore, India.

ABSTRACT

Every year, brands spend billions of dollars promoting their goods in stores. This marketing expenditure is scheduled three to six months in advance and is used to support promotional strategies that aim to improve brand awareness, encourage product and service trials, and boost consumption. COVID is upsetting this entire model. Because of the quick changes in consumer behavior, marketers can no longer afford to plan their promotional spending months. Brands must decide within a few days or even in close to real-time. Brands are moving their money to channels that are more adaptable as a result, such as digital advertising and promotions. Create a model that forecasts sales based on the sums spent on various marketing channels. With the contribution of advertising and social media marketing to generate sales, Advertisers are finding it more and more difficult to effectively measure the effectiveness of various revenue-generating marketing initiatives within their media mix as the advertising landscape continues to change. Our recommended system Creates a model that forecasts sales depending on the amount of money spent on various marketing platforms.

Keywords: linear regression, random forest, sale prediction, marketing channels.

1. INTRODUCTION

A straightforward Flask web application that uses a machine learning model to forecast sales based on spending on TV, radio, and newspaper advertising. Three sections make up the code. The first build two routes for the web app and imports the relevant libraries. It also initializes the Flask application. The "home" route gives the app an HTML template, whereas the "predict" route receives user input data, trains a machine learning model to anticipate sales, and then gives the app the prediction. A CSV file of advertising data is loaded into the code, and exploratory data analysis is carried out utilizing. It prepares the data for machine learning using the Seaborn and Matplotlib libraries. It divides the data into training and testing sets and trains the linear regression and random forest regression models, respectively. It shows the outcomes of the model's predictions against the actual sales data and assesses the models' performance based on mean squared error and R-squared. The Flask web app uses the joblib.dump() function to train the Random Forest Regression model, which is

used to forecast sales.

2. LITERATURE REVIEW

The literature has several studies that concentrate on estimating sales based on the volume of advertising spent.

The forecasting of media and advertising effects on sales is covered in reference [1] by Dawes et al. (2018), with a focus on the use of econometric models and other prediction techniques. This study examines the benefits and drawbacks of using time series analysis, structural equation modeling, and machine learning to forecast market outcomes.

A multiple linear regression method is presented in Reference [2] by Kologlu et al. (2018) for calculating the market worth of football players who play in the forward position. The study emphasizes the significance of looking at numerous player valuation aspects, such as performance metrics, age, experience, and contract status.

A description of multiple linear regression, a popular statistical method for predicting the connection between numerous independent factors and a dependent variable, is given in Reference [3] by Schinazi (2012). The paper discusses the method's presumptions and restrictions and techniques for choosing and interpreting models.

Reference [4] by Singh et al. (2017) focuses on a big data analytics investigation of Walmart's sales data. The research illustrates the difficulties in processing large and complicated datasets, including feature selection, model evaluation, and data preprocessing. The authors show how to anticipate sales results using a variety of machine learning methods, including decision trees, random forests, and support vector machines.

3. METHODOLOGY

To forecast sales based on the amount spent on advertising across various media channels (TV, radio, newspaper), the system was designed using a linear regression model and a random forest model. The system is built to accept as inputs the sums spent on advertising through different channels and produce a forecast sales value.

DATA COLLECTION:

A CSV file with data on the amount spent on sales and advertising was the source of the data used for the model

DATA EXPLORATION AND PREPROCESSING:

To find any trends or patterns in the data, visualization tools (such as seaborn and Matplotlib) were used to investigate the data. Additionally, the data was preprocessed by eliminating any null values and extra columns.

MODEL SELECTION:

A random forest model and a linear regression model were chosen as the two models for the system.

MODEL TRAINING AND TESTING:

The models were trained using the training data, and their performance was evaluated using the testing data.

MODEL EVALUATION:

Based on their mean squared error and r-squared values, the models were assessed.

DEPLOYMENT:

The joblib library was used to deploy the final model, which was the random forest model.   Overall, it appears that the development process is data-driven and iterative, with an placing a focus on data exploration and preprocessing, choosing and comparing various models, and implementing the best-performing model.

DATA SET

Each of the 200 cases in the dataset has four predictor variables: TV, Radio, Newspaper, and Sales. The information comprises the advertising expenditures made for a certain product on TV, radio, and newspapers as well as the related sales in thousands of units. The entire number of sales, expressed in thousands of units, is represented by the label variable "Sales". Each of the predictor variables is represented by 10 bins in the dataset, each of which contains a count of the instances that fall inside each range of values. TV, radio, and newspaper have respective ranges of 1.00 - 20.90, 0.70 - 30.27, and 0.00 - 4.96. The ranges for sales are 1.00 - 20.90, 20.90 - 40.80, 60.70 - 80.60, 80.60 - 100.50, 120.40 - 140.30, 160.20 - 180.10, and 180.10 - 200.00. The TV, Radio, Newspaper, and Sales variables' first 33 instances are represented by their precise values in the dataset, while the remaining values are represented by ellipses.

CLASSIFICATION METHOD

A regression issue where sales are forecasted based on the marketing budget. It involves loading and dividing the data, fitting linear regression and random forest regression models, making predictions on a test set, and measuring the models' performance using the mean squared error and R squared score. The code also saves the trained models using joblib and visualizes the model predictions and feature importances. Using the trained random forest regression model, it includes Flask code to build a web application that will enable users to input numbers for TV, radio, and newspaper advertising spending and get a forecast sales value. Regression with web application development is the classification for this code.

In machine learning, different models are further divided into classification approaches:

1. Linear Regression

2. Random Forest

It specifically uses two different models, linear regression and random forest regression, to forecast sales depending on the amount spent on advertising. The code reads data from a CSV file, divides it into training and testing sets, fits the two regression models, assesses them using mean squared error and R-squared score, shows the outcomes, and then does the same for the other regression model. The value of a feature. Finally, joblib is used to save the learned models. Additionally, the program contains a Flask web application that enables users to input newspaper, radio, and television advertising budgets and obtain a sales forecast using the learned models.

LINEAR REGRESSION

Using the sci-kit-learn toolkit, the linear Regression model is trained to forecast sales based on the amount spent on advertising. The "Advertising.csv" CSV file is used to load the dataset used to train the model. Using the train_test_split, Training sets, and testing sets are created from the data. Then, using the fit method, a Linear

Regression model is built using the Linear Regression class from sci-kit-learn and fitted to the training set of data. After the model has been trained and applied to making predictions on the testing data using the predict method, the mean squared error and R-squared score of the model are calculated using the mean_squared_error and r2_score functions from sci-kit-learn. To plot the expected vs. actual sales and the predicted vs. real sales, a scatter diagram of the findings from the linear regression model is employed.

RANDOM FOREST

The use of the Random Forest Regressor for sales forecasting based on advertising expenditure. Several decision trees are built using the Random Forest machine learning technique, and their predictions are then combined to give the outcome.

After loading the data and separating it into training and test sets, a Random Forest Regressor is fitted with 100 trees on the training set and the training set. The model's performance is then assessed by predicting the sales for the testing set and using the mean squared error and R-squared score. The most crucial predictors are identified by plotting the feature importances of the Random Forest model. Finally, it uses the joblib library to save the trained Random Forest model to a file for later use.

4. RESULT AND DISCUSSION

To forecast the outcome, a particular dataset and user input are needed. The code offers two distinct models for predicting sales based on information about advertising expenditures: a linear regression model and a random forest regression model. Plotting the outcomes and preserving the trained models are also included. The code for the Flask web application is made to load the stored model and utilize it to create predictions depending on input from the user for advertising spending numbers. The outcome would be determined by the precise input values that the user-supplied.

Fig1 Prediction output
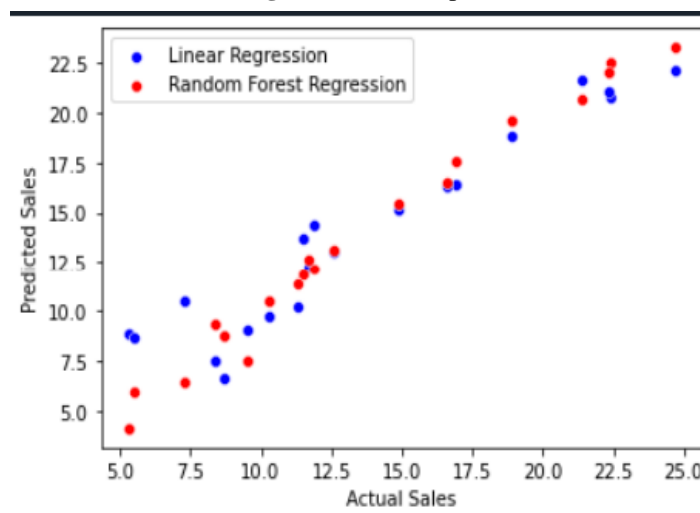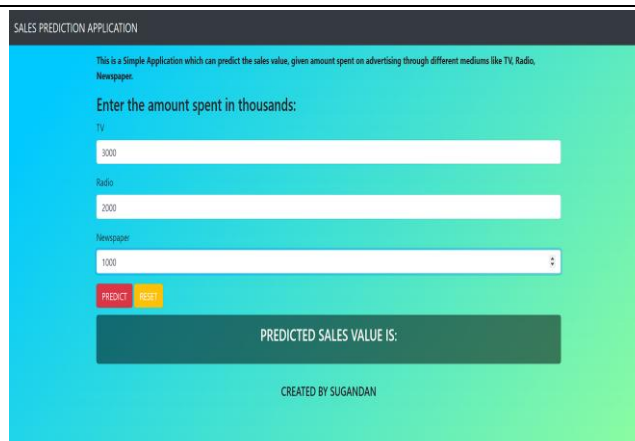
Fig2 Scatter plot of the used algorithm
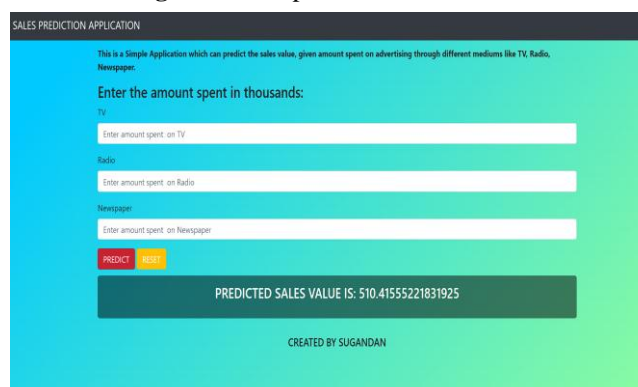
Fig3 Amount spend on advertisement

Fig4 Predicted sales value

5. CONCLUSION

on social media Modern company depends heavily on advertising. It helps marketers strengthen relationships with customers and increase sales. Marketers utilize social media to advertise their products and boost sales. The study seeks to forecast how much advertising spending will affect a company's sales. The goal is to develop a model that forecasts sales using advertising spending on television, radio, and newspapers. The open-source Python Anaconda statistical analysis software is used to analyze the data. Python is a potent programming language that uses a variety of packages and may graphically display a dataset about various factors. The findings indicate that television advertising outperforms radio and newspaper marketing as a predictor of business sales.

6. REFERENCES

[1] Dawes, John & Kennedy, Rachel & Green, Kesten & Sharp, Byron. (2018). Pre-publication version to Forecasting advertising and media effects on sales: Econometrics and alternatives. International Journal of Market Research. 60.10.1177/1470785318782871.

[2] Kologlu, Yunus & Birinci, Hasan & Ilgaz, Sevde & Ozyilmaz, Burhan. (2018). A Multiple Linear Regression Approach For Estimating the Market Value of Football Players in Forward Position.

[3] Schinazi, Rinaldo. (2012). Multiple Linear Regression. 10.1007/978-0-8176-8250-7_10.

[4] Singh, Manpreet & Ghutla, Bhawick & Jnr, Reuben & Mohammed, Aesaan & Rashid, Mahmood. (2017). Walmart's Sales Data Analysis - A Big Data Analytics Perspective. 114-119. 10.1109/APWConCSE.2017.00028.

[5] Gonzalo, Frederic, "The Impact of Social Media in Ad Spending [Infographic]." Frederic In 2018, Gonzalo's Blog.

[6] Michael Lewis. The phrase "How to Use Social Media Marketing Strategy to Increase Sales." Cash Grabbers.

[7] Dennis, C., Giglo, S., and Pentano, E. (2018). Understanding consumer sentiment from their tweets: Big data analytics for quick fashion businesses. Journal of Retail and Distribution Management International.

[8] Vasista, T.G., Zamil, A.M.A., & 2020. Realizing RFM analysis customer segmentation using Python implementation. Indian Journal of Marketing, Prabandhan (submitted in July 2020).