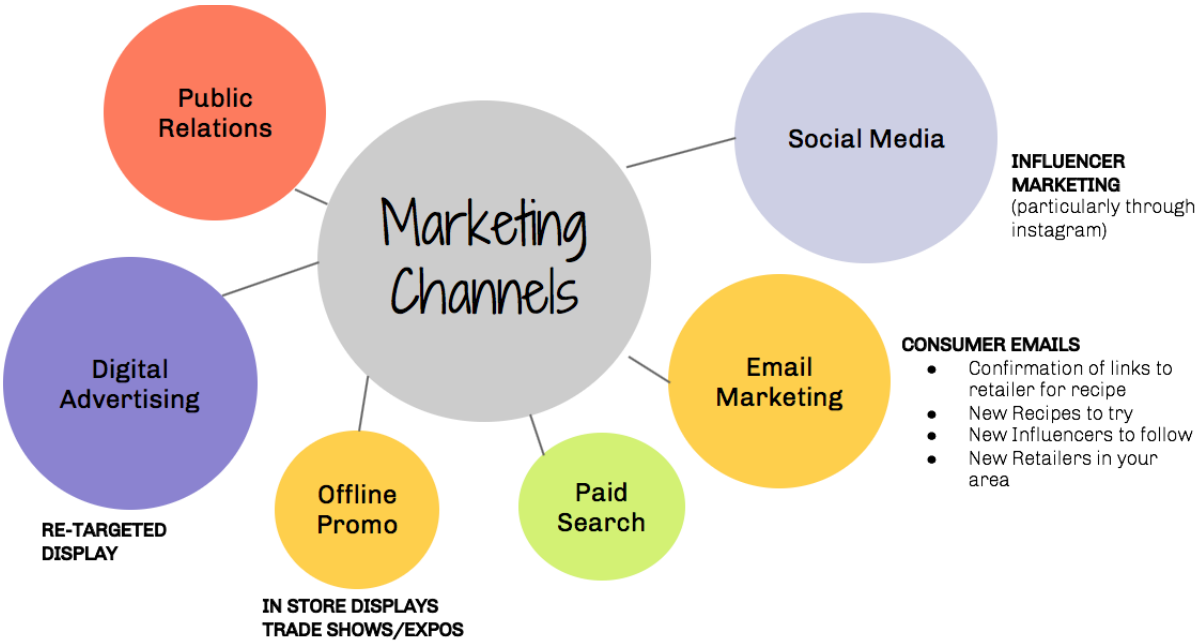
ADVERTISING SALES CHANNEL PREDICTION MODEL



Submitted by:

UJJWAL PRATIK

Contents

ACKNOWLEDGMENT

I sincerely thanks to the Data Trained Faculty for the guidance. They have covered the topics like Machine Language, Python & SQL. I had also taken help from YouTube & online videos.

INTRODUCTION

ABOUT ADVERTISEMNT

- The aim of the marketing strategy is to use advertising as well as other tools to maximize sales.
- Advertising is a tool which can attract the prospective customers to a product or services.
- Advertising strategies are part of overall marketing strategies which also include public relations, incentives, word of mouth etc.
- Advertising stimulates demand, develops brand preference, cuts cost, and acts as a competitive weapon.
- It involves an adequate mix of various types of advertising like, competitive, persuasive, informative to gain attention of consumers and initiate purchase.
- Companies rely on these strategies to persuade consumers to buy their products.

PROBLEM STATEMENT

- When a company enters a market, the distribution strategy and channel it uses are keys to its success in the market, as well as market know-how and customer knowledge and understanding. Because an effective distribution strategy under efficient supply-chain management opens doors for attaining competitive advantage and strong brand equity in the market, it is a component of the marketing mix that cannot be ignored.
- ds. The distribution strategy and the channel design must be right the first time. The case study of Sales channel includes the detailed study of TV, radio, and newspaper channel. The predict the total sales generated from all the sales channels.

OBJECTIVE

- Our main objective is to predict the total sales using machine learning algorithms.
- All the parameters will be analysed through Machine Learning algorithms like Linear Regression, Lasso and Ridge Regression which will help to predict the total sales generated from all the sales channels.

DATA DESCRIPTION

- The dataset contains the detailed study of TV, radio, and newspaper channel.
- The source of data is taken from GitHub.

DATA SAMPLE LOOK USING PANDA

	Unnamed: 0	TV	radio	newspaper	sales
0	1	230.1	37.8	69.2	22.1
1	2	44.5	39.3	45.1	10.4
2	3	17.2	45.9	69.3	9.3
3	4	151.5	41.3	58.5	18.5
4	5	180.8	10.8	58.4	12.9

METHEDOLOGY

- It gives insights of the dependency of target variables on independent variables using machine learnings techniques to determine the total sales because it gives the best outcome for the assurance of sales generated.
- The dependent variable is sales, whereas other variables i.e., Tv, radio and newspaper etc. are independent variables.

METRIC USAGE

a. Linear Regression.
b. Lasso Regression.
c. Ridge Regression.
d. Elastic Net Regression.
e. Support Vector Regression.
f. Random Forest Regression.

Hardware and Software Requirements and Tools Used

a. Hardware Requirement:
    i. Intel core i5
    ii. 8 GB Ram
b. Software Requirement:
    i. Python 3.x with packages:
        1. Pandas: Data analysis and manipulation tool
        2. NumPy: Provide support for mathematical functions, random number etc.
        3. Matplotlib: is a low-level graph plotting library in python that serves as a visualization.
        4. Seaborn: is a library mostly used for statistical plotting in python.
        5. Scikit-Learn: is an open-source Python library that has powerful tools for data analysis and data mining.

IDENTIFICATION OF POSSIBLE PROBLEM-SOLVING APPROACHES

Following models are used for solving the problem:

a. R2 score: is used to evaluate the performance of a linear regression model.
b. Linear Regression: Logistic regression is fast and relatively uncomplicated, and it is convenient for you to interpret the results.
c. Lasso: The Lasso is a linear model that estimates sparse coefficients with l1 regularization.
d. Ridge: Ridge regression is an extension of linear regression where the loss function is modified to minimize the complexity of the model.
e. Elastic Net: is a linear regression model trained with both l1 and l2 -norm regularization of the coefficients.
f. Support Vector Regressor: Support Vector Regression as the name suggests is a regression algorithm that supports both linear and non-linear regressions. This method works on the principle of the Support Vector Machine.
g. Cross-Validation-Score: a model that would just repeat the labels of the samples that it has just seen would have a perfect score but would fail to predict anything useful on yet-unseen data.
h. Random Forest Regressor: is a meta estimator that fits several classifying decision tree on various sub samples of the dataset.
i. Grid Search CV: This function helps to loop through predefined hyperparameters and fit your estimator (model) on your training set.
j. Mean Squared Error: this metric gives an indication of how good a model fits a given dataset.
k. Root Mean Squared error : is a frequently used measure of the differences between values (sample or population values) predicted by a model or an estimator and the values observed.

TESTING OF IDENTIFIED APPROACH(Algorithms)

a. Train Test Split
b. Linear Regression
c. Lasso Regression
d. Ridge Regression
e. Elastic Net Regression
f. Support Vector Regression
g. Random Forest Regression
h. Grid Search CV
i. Cross Validation
j. Mean Squared Error
k. Root Mean Squared Error
l. R2 score

KEY METRICS FOR SUCCESS IN SOLVING PROBLEM UNDER CONSIDERATION.

1. Analysed data for any outliers and removed it by z-score method.
2. Analysed data for any skewness.
3. Cross Validation for cross validates the accuracy-score from overfitting.
4. Hyper parameter tuning using Grid Search CV, Random Forest Regressor for making the prediction better.

CONCLUSION

- Companies engage in advertising of their products to increase awareness and induce purchase.
- The various mix of advertising strategies have enhanced sales and has paved way for the product's success in the competitive market.

APPLICATIONS

- Results will be used by companies to improve the sales of the future.

REFERENCES

- Data trained course videos.
- Google Search.
- YouTube.
- GitHub.
- UCI Machine learning repository.

LEARNING OUTCOMES OF THE STUDY IN RESPECT OF DATA SCIENCE

This study gives me opportunity for lots of learning starting from various types of plotting like histograms, boxplot, scatterplot, line chart and many more graphs. These graphs helped me to analyse different aspects of data like outlier, skewness, correlation etc.

It also helped me to learn how to apply various model techniques on data and enable predications.