TASK4 - SALES PREDICTION

CODSOFT INTERNSHIP23

Problem Statement

Sales prediction involves forecasting the amount of a product that customers will purchase, taking into account various factors such as advertising expenditure, target audience segmentation, and advertising platform selection.

Objective

Build a model which predicts sales based on the money spent on different platforms for marketing.

Project Approach

- Framing Questions: This is the very first step of the analysis process. Going through the problem statement deeply and extract some hidden questions that problem statement is throwing to us is the main objective here.

- Data Inspection: This is the second step involved here which includes checking dataset information about its features and structure, handling null/missing values (if any) which is basically data cleaning.

- Data Pre-processing: This is the third step about feature overview where segregating features into dependent (target variable) and independent features (input variables).

- EDA: Exploratory Data Analysis (EDA) takes the fourth step where the data visualization methods are used to check distribution of all the features in the data and also to check different insights about dataset.

- Model Implementation: Fifth step is all about implementing the model which involves splitting data into train and test sets and fitting different ml algorithms for the purpose of prediction of target variable or output variable.

- Model Evaluation: This step is all about comparing evaluation metrics of each of the models as to find out the best fit model for this particular case.

- **Conclusions:** Last step is all about summing up information and performance of the model observed from the analysis process.

Algorithms:

In this case three different machine learning algorithms has been taken into consideration. They are:

➢ **Linear Regression:** It is one of the most popular and easiest ml algorithms. In statistics, linear regression is a statistical model which estimates the linear relationship between a scalar response and one or more explanatory variables (also known as dependent and independent variables).

➢ **Ridge Regression:** Ridge regression is a method of estimating the coefficients of multiple-regression models in scenarios where the independent variables are highly correlated. This is also called as L2 regularization.

➢ **Lasso Regression:** In statistics and machine learning, lasso (least absolute shrinkage and selection operator; also Lasso or LASSO) is a regression analysis method that performs both variable selection and regularization in order to enhance the prediction accuracy and interpretability of the resulting statistical model. The lasso method assumes that the coefficients of the linear model are sparse, meaning that few of them are non-zero.

Final Conclusion:

Going through all the predictive analysis process, it was concluded that the 'TV' which is the independent variable contributing more towards 'sales' (output variable) as compared to other input variables.