APPLIED DATA SCIENCE GROUP-2

FUTURE SALES PREDICTION (PHASE -2)

WHAT IS FUTURE SALES PREDICTION?

Future sales prediction refers to the process of using historical sales data, market trends, and various analytical techniques to forecast or estimate the future sales performance of a product. It involves using mathematical models, sometimes machine learning algorithms to predict the sales.

These predictions are valuable for businesses as they can help in making informed decisions regarding the accurate sales . By accurately predicting future sales, companies can optimize their operations, meet customer demand more effectively.

The accuracy of future sales predictions can vary depending on the quality of data.

One of the most common methods used to predict sales is regression analysis. This method involves using historical sales data to train a model that can predict future sales. The model can take into account factors such as past sales, marketing campaigns, and economic indicators to make its predictions.

DATASETS:

Dataset: www.kaggle.com

Dataset name: future-sales-prediction

Dataset link: https://www.kaggle.com/datasets/chakradharmattapalli/future-sales-prediction

DETAILS ABOUT THE COLUMNS:

It consist of 4 columns and every column consist of 200 data's

TV

RADIO

NEWSPAPER

SALES

LIBRARIES:

PACKAGE INSTALLATION (Note: All the packages are installed through command prompt)

Pandas is a powerful library for data manipulation and analysis. It provides data structures like DataFrames for working with structured data, such as tabular data in CSV or Excel files.

INSTALLATION- pip install pandas

Numpy is used for numerical computations and working with arrays.

INSTALLATION- pip install numpy

Matplotlib is a popular library for creating static, animated, or interactive visualizations in Python. The pyplot module provides a simple interface for creating various types of plots and charts.

INSTALLATION- pip install matplotlib

Seaborn is built on top of matplotlib and provides a high-level interface for creating attractive and informative statistical graphics.

INSTALLATION- pip install seaborn

This line imports the `train_test_split` function from the `model_selection` module of the scikit-learn library. It's used to split the dataset into training and testing sets.

INSTALLATION- pip install scikit-learn

HOW TO TEST AND TRAIN:

Scikit-learn alias **sklearn** is the most useful and robust library for machine learning in Python. The **scikit-learn library** provides us with the model_selection module in which we have the splitter function train_test_split().

Syntax - from sklearn.model_selection import train_test_split

Load your dataset and prepare it for splitting. This typically involves loading the data, preprocessing it (e.g., handling missing values, encoding categorical variables), and separating the features (X) and the target variable (y). Use the train_test_split function to split your data into training and testing sets. Specify the proportion of data you want to allocate to the training set using the test_size(test_size=0.2)

80% training set

20% testing set

Use the training data (X_train and y_train) to train your machine learning model.

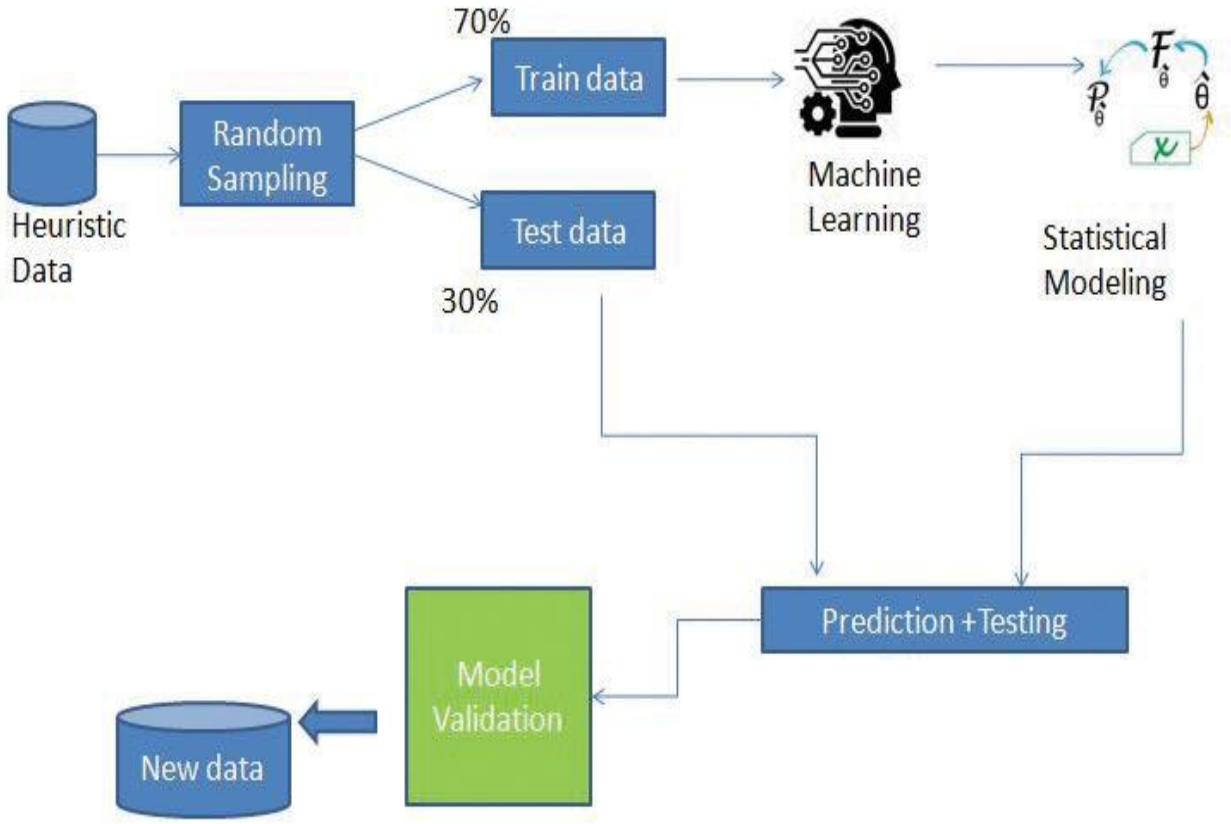
Train your model:

```
from sklearn.linear_model import LinearRegression
model = LinearRegression()
model.fit(X_train, y_train)
```

Evaluate Your Model:

```
y_pred = model.predict(X_test)
```

PROCESS FLOW:



LINEAR REGRESSION:

Linear regression is one of the easiest and most popular Machine Learning algorithms. It is a statistical method that is used for predictive analysis. Linear regression is a supervised learning algorithm that compares input (X) and output (Y) variables based on labeled data. It's used for finding the relationship between the two variables and predicting future results based on past relationships.