GRT INSTITUTE OF ENGINEERING AND TECHNOLOGY, Tiruttani

1103-GRT INSTITUTE OF ENGINEERING AND TECHNOLOGY

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

PROJECT TITLE

Future sales prediction

Harini M
3rd year, 5th sem
Reg no.:110321104013
harinimani30112004@gmail.com

INNOVATION

2.1 FUTURE SALES PREDICTION

Future sales prediction, also known as sales forecasting, is the process of estimating a company's future sales based on historical data, market trends, and various analytical techniques. Accurate sales predictions are crucial for businesses as they help in making informed decisions regarding inventory management, resource allocation, budget planning, and overall strategy. Here's an explanation of the key aspects of future sales prediction:

- Historical Data Analysis
- Market Research
- Quantitative Method

1. Time Series Analysis
2. Regression Analysis
3. Machine Learning and AI

- Qualitative Factors
- Scenario Analysis
- Technology and Software
- Continuous Monitoring and Updating
- Feedback Loop
- Collaboration

2.2 DATASET

This dataset is designed for forecasting future sales in a retail context. It contains historical sales data along with various features that can be used to build predictive models.

When working with a dataset, it's essential to perform data exploration and preprocessing to better understand its characteristics and prepare it for modeling.

We've got the data set in the website called Kaggle(www.kaggle.com/data)

The data set which is respective to our project is sales.csv

(**https://www.kaggle.com/datasets/chakradharmattapalli/future-sales-prediction**)

The data set having the 4columns named TV,Radio,Newspaper,Sales and having 200 rows of datas.(numerical values).

The dataset given here contains the data about the sales of the product. The dataset is about the advertising cost incurred by the business on various advertising platforms. Below is the description of all the columns in the dataset:

TV: Advertising cost spent in dollars for advertising on TV;

Radio: Advertising cost spent in dollars for advertising on Radio;

Newspaper: Advertising cost spent in dollars for advertising on Newspaper;

Sales: Number of units sold

So, in the above dataset, the sales of the product depend on the advertisement cost of the product.

2.3DETAILS ABOUT COLUMNS

To predict future sales, you can create a dataset with columns for TV advertising spending, radio advertising spending, newspaper advertising spending, and sales. These columns will help you build a regression model to predict sales based on advertising expenditures. Here are some details about each of these columns:

TV:

Column Name: TV

Data Type: Numeric (continuous)

Description:

This column represents the amount of money spent on advertising through television channels. It includes expenses for television commercials, sponsorships, and other TV-related advertising efforts. Measured in dollars.

Radio:

Column Name: Radio

Data Type: Numeric (continuous)

Description:

This column represents the amount of money spent on advertising through radio channels. It includes expenses for radio commercials, radio show sponsorships, and other radio-related advertising efforts. Measured in dollars.

Newspaper:

Column Name: Newspaper

Data Type: Numeric (continuous)

Description:

This column represents the amount of money spent on advertising in newspapers. It includes expenses for print advertisements, classified ads, and other newspaper-related advertising efforts. Measured in dollars.

Sales:

Column Name: Sales

Data Type: Numeric (continuous)

Description:

This is the target column you want to predict. It represents the actual sales revenue generated as a result of the advertising expenditures on TV, radio, and newspaper. Measured in dollars.

TV	Radio	Newspaper	Sales
203.1	37.8	69.2	22.1
44.5	39.3	45.1	10.4
17.2	45.9	69.3	12
151.5	41.3	58.5	16.5
180.8	10.8	58.4	17.9

You can collect historical data for these columns and then use regression techniques (e.g., linear regression) to build a model that predicts future sales based on advertising expenditures. Additionally, you may want to include other relevant features or variables, such as seasonality, market trends, or competitor advertising spending, to improve the accuracy of your sales prediction model.

2.4LIBRARIES

To work with the data and build a predictive model for future sales based on the TV, radio, and newspaper advertising expenditures, you will typically use Python and several libraries.

Here are the key libraries you may need and how to download them: make sure to activate your virtual environment first and then run the pip install commands within that environment.

- Numpy
- Pandas
- Matplotlib and seaborn
- Sci-kit Learn(sk learn)

Import Libraries

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, r2_score
```

Load the data

```
df = pd.read_csv('your_dataset.csv')
```

Testing for null values

```
#So ,this dataset doesn't have any null values.
print(data.isnull().sum())
```

Model Building

```
X = df[['TV', 'Radio', 'Newspaper']]
y = df['Sales']
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
#for future sales prediction
model = LinearRegression()
model.fit(X_train, y_train)
model.fit(X_train, y_train)
#features = [[TV, Radio, Newspaper]]
features = np.array([[230.1, 37.8, 69.2]])
print(model.predict(features))
```

Model Evaluation

```
y_pred = model.predict(X_test)
mse = mean_squared_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)
```

2.5TRAIN AND TEST

Code

```
x = np.array(data.drop(["Sales"], 1))
y = np.array(data["Sales"])
xtrain, xtest, ytrain, ytest = train_test_split(x, y,  test_size=0.2, random_state=42)
```

Visualization

```
#relationship between the amount spent on advertising on TV and units sold
import plotly.express as px
import plotly.graph_objects as go
figure = px.scatter(data_frame = data, x="Sales",y="TV", size="TV", trendline="ols")
figure.show()
```

2.6 REST OF EXPLANATION

Training and testing data for future sales analysis typically involves using historical sales data to build and evaluate predictive models. These models can help you forecast future sales, identify trends, and make informed business decisions. Here's a step-by-step guide on how to train and test data for future sales analysis. Split your dataset into two parts: a training set and a testing set.

A common split is 70-80% for training and 20-30% for testing. You can also use time-based splitting where the training data comes from earlier time periods, and the testing data comes from more recent periods.

2.7 METRICS USED FOR ACCURACY CHECK

Several metrics are commonly used to check the accuracy of predictive models, depending on the specific problem and the nature of the data. Here are some of the most commonly used metrics for accuracy checking:

- Mean Absolute Error (MAE)
- Mean Squared Error (MSE)
- Root Mean Squared Error (RMSE)
- Mean Absolute Percentage Error (MAPE)
- R-squared (R2)
- Adjusted R-squared
- Accuracy, Precision, Recall, F1-Score (Classification Metrics)
- Area Under the Receiver Operating Characteristic Curve (AUC-ROC)
- Confusion Matrix
- Log-Loss (Logarithmic Loss)
- Cohen's Kappa

The choice of metric depends on the problem you're solving and the nature of your data. For regression problems, metrics like MAE, MSE, and RMSE are common, while classification problems typically use accuracy, precision, recall, F1-score, AUC-ROC, and log-loss.

It's essential to select the most appropriate metric based on the specific goals and characteristics of your analysis.

SUMMARY

So this is how we can train a machine learning model to predict the future sales of a product. Predicting the future sales of a product helps a business manage the manufacturing and advertising cost of the product.

In essence, future sales prediction is a dynamic and data-driven process, combining historical insights with forward-looking analysis to empower businesses to make proactive decisions and thrive in an ever-evolving marketplace. Accurate sales predictions enable organizations to streamline operations, allocate resources effectively, and remain agile in the face of market fluctuations.