Future sales prediction

Team members:

Srudhi A
Sushmitha S
Saleeka Barwin P
Aswathy Asokan
Thanisha.T

PHASE 4 PROJECT SUBMISSION


%
$
2023

Dataset

- A dataset for future sales prediction typically contains historical sales data, along with various relevant features or attributes that can help in forecasting future sales. This type of dataset is commonly used in retail, e-commerce, and other businesses to make informed decisions about inventory management, marketing strategies, and overall business planning.
- Here's a detailed explanation of the components of a future sales prediction dataset
- **Date or Time Period**: This field records the date or time period of each sales transaction or observation. It can be broken down into years, months, days, or even hours, depending on the level of granularity required for the analysis.

- **Sales Amount**: The most crucial field, which records the actual sales amount for each transaction. This can be measured in terms of currency (e.g., dollars, euros) and is the target variable for prediction.
- **Product ID or SKU**: An identifier for the product being sold. This can be used to categorize sales data and analyze the performance of specific products.
- **Store or Location ID**: If sales data involves multiple stores or locations, this field helps in tracking sales by store, which can be useful for store-level predictions

- **Promotions and Discounts**: Information about any promotions, discounts, or special offers that were in effect during the transaction. This can help in understanding the impact of marketing activities on sales.
- **Weather Data**: In some cases, weather conditions at the time of the transaction might be included. This is especially relevant for businesses that are weather-dependent, like clothing retailers.
- **Holiday or Seasonal Data**: Details about holidays and seasons can be included to account for seasonality in sales. For example, Christmas, Black Friday, or back-to-school season.
- **Customer Information**: If applicable, data about the customer making the purchase, such as demographics or loyalty status. This can help in customer segmentation and personalized marketing.

- The goal of using such a dataset is to build a predictive model that can forecast future sales based on historical data and these relevant features. Various machine learning and statistical techniques can be applied, such as time series analysis, regression models, and deep learning methods, to create accurate predictions. These predictions can then be used for demand forecasting, optimizing inventory, marketing strategies, and overall business planning.
- Keep in mind that the quality and quantity of data, as well as the choice of modeling techniques, play a significant role in the accuracy of future sales predictions. Additionally, continuous data collection and model retraining are often necessary to adapt to changing market conditions.

Preprocessing

- Preprocessing a dataset for future sales prediction is a critical step in the data analysis and modeling process. Proper data preprocessing can significantly impact the accuracy and reliability of your predictive models. Here's a step-by-step guide on how to preprocess a dataset for future sales prediction:
- **Data Encoding**:
- Convert categorical variables into numerical format using techniques like one-hot encoding or label encoding.
- Scale or normalize numerical features to ensure they are on the same scale. Common methods include Min-Max scaling and standardization.

Codings

```
import pandas as pd
import numpy as np
data = pd.read_csv("sales_data.csv")

data = data.drop_duplicates()
data = data.dropna()
selected_features = data[['Date',
'SalesAmount', 'ProductID', 'Promotions']]
data_encoded = pd.
get_dummies(selected_features,
columns=['ProductID'])
from sklearn.model_selection import
train_test_split
X = data_encoded.drop('SalesAmount',
axis=1)
```

```
y = data_encoded['SalesAmount']
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
from sklearn.preprocessing import MinMaxScaler
scaler = MinMaxScaler()
X_train[['Promotions']] = scaler.fit_transform(X_train[['Promotions']])
X_test[['Promotions']] = scaler.transform(X_test[['Promotions'])
```

Conclusion

- In conclusion, predicting future sales is a dynamic and multifaceted process. By leveraging high-quality data, advanced analytics, an understanding of market dynamics, and ongoing collaboration and adaptation, businesses can make more accurate predictions and position themselves for success in the future. Remember that while predictions can provide valuable insights, they should be used in conjunction with sound business judgment and a flexible approach to adapt to changing circumstances.

Thank you