Sales prediction for a retail store

Harsha c k

Harsha C K

I have completed BSC mathematics and Currently I am Pursuing Masters Certificate programme in Artificial Intelligence , Sree Shankaracharya Computer Centre and Pragyan AI .

Introduction

What is Sales Prediction?
Sales prediction involves using historical data and statistical/machine learning models to forecast future sales. It helps businesses optimize inventory, marketing strategies, and financial planning.

Importance in Retail:

- Helps retailers prevent stock shortages and overstocking.
- Aids in strategic pricing and promotion planning.
- Enhances revenue forecasting for better business decisions.

Problem Statement

Challenges in Retail Sales Forecasting:

- **Fluctuations in Demand:** Seasonal trends, holidays, and promotions cause unpredictable demand changes.
- **External Factors:** Weather, economy, and social trends impact sales.
- **Data Limitations:** Missing or unstructured data can lead to inaccurate predictions.

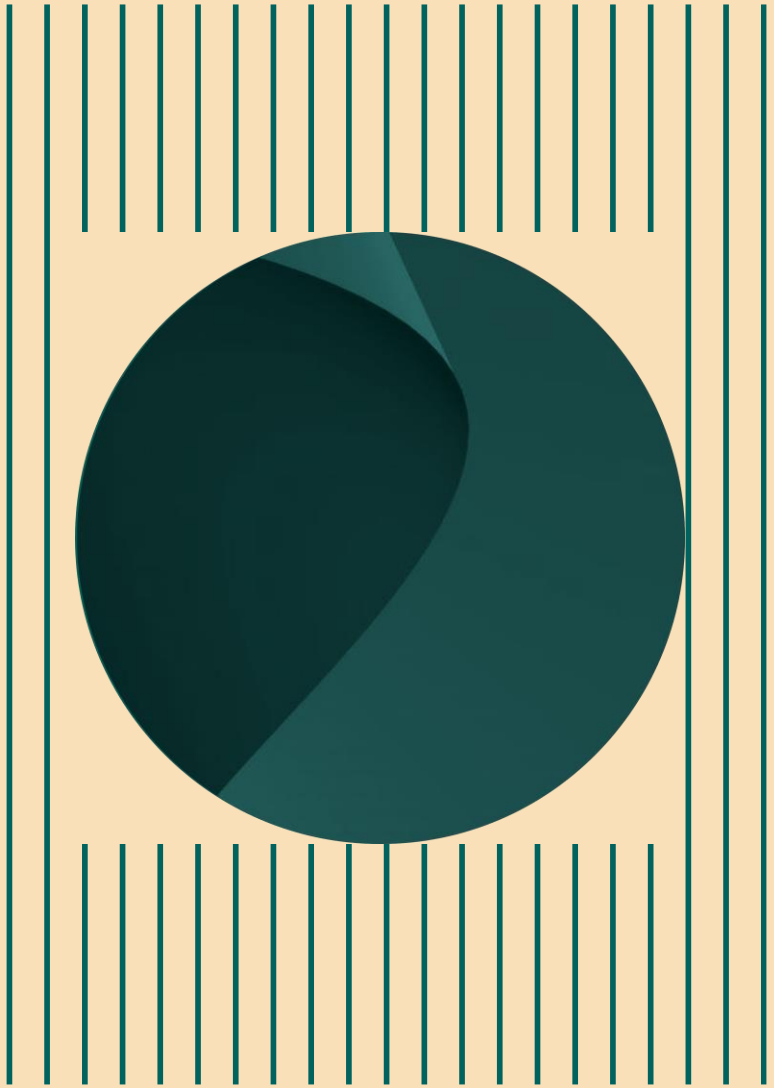
Dataset & Features

Data Sources:

- Historical sales records
- Customer transactions
- Weather and seasonal patterns
- Promotions and marketing campaigns

Key Features in Sales Prediction:

- **Date:** Helps identify daily, weekly, or seasonal trends.
- **Product Category:** Different items have different demand patterns.
- **Price:** Sales may fluctuate based on pricing and discounts.
- **Promotions & Discounts:** Special deals often increase sales.
- **Weather Conditions:** Weather can influence purchasing behavior (e.g., rain affecting store footfall).
- **Customer Footfall:** The number of customers visiting the store impacts total sales.



Methodology

Step 1: Data Collection & Preprocessing

- Gather sales data from databases or spreadsheets.
- Handle missing values by filling in or removing incomplete records.
- Normalize or scale data if necessary (especially for price and sales figures).

Step 2: Exploratory Data Analysis (EDA)

- Use visualizations (line charts, histograms) to detect sales patterns.
- Identify correlations between variables (e.g., discounts vs. sales increase).

Step 3: Feature Engineering

- Create new features such as "holiday season flag" or "discount percentage."
- Convert categorical data (e.g., product category) into numerical form using encoding techniques.

Step 4: Model Selection

- Choose the best model based on data type and prediction needs.
- Implement different algorithms such as regression, decision trees, or time series forecasting.

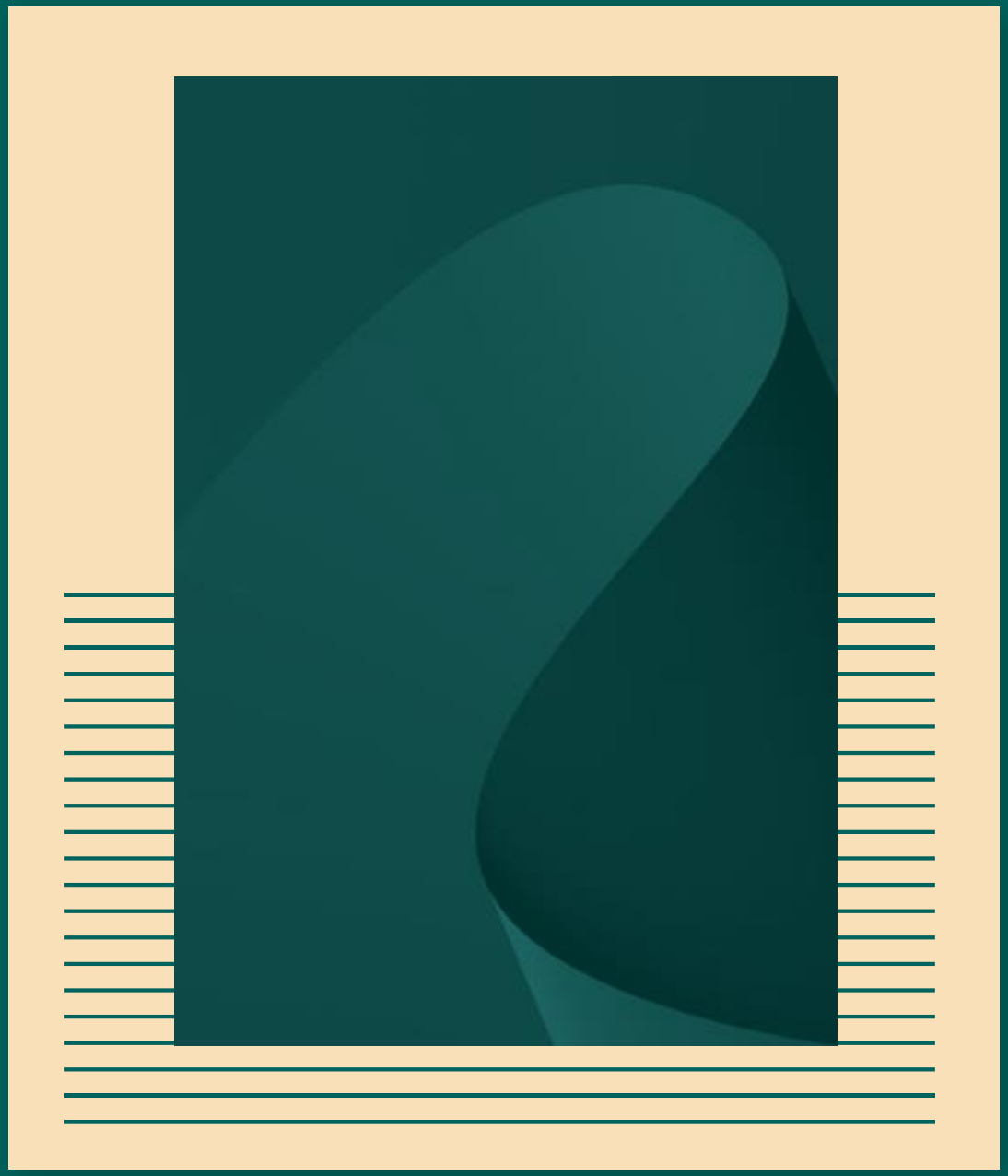
Step 5: Model Evaluation

- Use metrics like Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and R^2 Score to assess model performance.
- Compare models to select the most accurate one.

Model Selection

Types of Models Used:

- **Linear Regression:** Simple and interpretable; best for basic sales trends.
- **Decision Trees & Random Forest:** Handles complex, non-linear relationships well.
- **XGBoost:** Powerful model that performs well on structured tabular data.
- **LSTM (Long Short-Term Memory):** A deep learning model ideal for time series forecasting.



Model Implementation

We train a **Random Forest Regressor**, a machine learning model that predicts sales based on multiple input features.

Results & Evaluation

Model Performance Metrics:

- **RMSE (Root Mean Square Error):** Measures how far predictions deviate from actual sales.
- **MAE (Mean Absolute Error):** Shows average absolute error between predicted and actual sales.
- **R^2 Score:** Indicates how well the model explains sales variability.

Comparison of Different Models:Best performing model based on highest accuracy.

Conclusion & Future Scope

Key Takeaways:

- Machine learning models improve sales forecasting accuracy.
- Using external data (e.g., weather, holidays) enhances predictions.
- Feature engineering is crucial for improving model performance.

Future Improvements:

- **Incorporating Social Media Trends:** Customer sentiment analysis for better predictions.
- **AI-driven Demand Forecasting:** Using deep learning models like Transformer-based networks.
- **Real-Time Prediction Systems:** Deploying models in production for dynamic forecasting.

Thank You

Harsha C K

Github : https://github.com/harshakadakam/Huber-Regression-case-study-project.git

Linkedin : https://www.linkedin.com/in/harsha-kadakam-a923b8347