Sales Prediction

Project Overview

This project focuses on building a **machine learning model** to **forecast product sales** based on historical data and marketing factors like advertising spend, promotions, and customer segmentation.
The goal is to help businesses optimize their marketing strategies and boost overall sales growth.

Dataset

- **Source**: [Uploaded locally]
- **Fields Include**:
  - `Date` — time of the sale.
  - `Product Category` — product type.
  - `Sales` — units sold (target variable).
  - `Advertising Spend` — marketing budget.
  - `Promotion` — discount or promotional offers.
  - `Customer Segment` — type of customers.
  - (Other business-specific features.)

Project Steps

1. Data Loading

- Imported the dataset using the **Var. File** node in SPSS Modeler.

2. Data Understanding

- Analyzed the structure and distributions using **Table**, **Plot**, and **Data Audit** nodes.

- Identified missing values, outliers, and feature types.

3. Data Preparation

- **Missing Values**: Handled using:
  - Imputation (mean/mode) for numerical/categorical fields.
- **Outliers**: Detected via boxplots and corrected using capping techniques.
- **Feature Scaling**: Applied Z-score normalization for numeric fields like `Advertising Spend` and `Sales`.
- **Feature Engineering**: Created new features like `Month`, `Season`, and interaction terms.

4. Data Partitioning

- Divided the dataset:
  - **Training Set** (70%)
  - **Testing Set** (30%)
- Used the **Partition** node to ensure unbiased evaluation.

5. Model Building

- Built and compared multiple predictive models:
  - **Linear Regression**
  - **Decision Trees (CHAID, CART)**
  - **Random Forest Regression**
- Target variable: `Sales`
- Input features: All other relevant fields.

6. Model Evaluation

- Evaluated models based on:
  - **R-squared (R^2)**
  - **Root Mean Square Error (RMSE)**
  - **Mean Absolute Error (MAE)**
- Selected the model with the best generalization performance on the test set.

7. Model Deployment

- Deployed the best model using the **Apply Model** node to predict future sales.

📈 Results

Metric	Best Model (Example)
R^2	0.87
RMSE	1200
MAE	900

(Values vary depending on data preprocessing.)

🛠️ Tools Used

- **IBM SPSS Modeler 18.x**
- **Python (optional preprocessing)**
- **GitHub** for version control and documentation.

📂 Project Structure

```
arduino
CopyEdit
Sales_Prediction/
├── data/
│   ├── sales_data.csv
│   └── processed_sales_data.csv
├── models/
│   └── sales_prediction_model.str (SPSS Stream)
├── outputs/
│   ├── evaluation_metrics.png
│   └── sales_predictions.csv
```

```
├── README.md
└── report/
    └── Sales_Prediction_Report.pdf
```

🚀 How to Run

1. Open **IBM SPSS Modeler**.
2. Load the `.str` file from the `models/` folder.
3. Link the `sales_data.csv` file properly.
4. Run the stream to train and evaluate the models.
5. Check evaluation outputs in the outputs folder.

📬 Contact

For questions or contributions, feel free to reach out!

☑ Project Completed!

🚀