

## # Future Sales Prediction README

This README file provides detailed information on running the Future Sales Prediction code, its dependencies, and an explanation of the code.

### ## Table of Contents

1. [Introduction](#introduction)
2. [Dataset Description](#dataset)
3. [Dependencies](#dependencies)
4. [How to Run the Code](#how-to-run-the-code)
5. [Code Explanation](#code-explanation)

### ## Introduction

The Future Sales Prediction code is designed to predict sales based on advertising spending data. It uses various machine learning models, including Linear Regression, Random Forest, Support Vector Machine, Neural Network, XGBoost, and LightGBM. Additionally, the code includes time series forecasting using ARIMA, ETS (Exponential Smoothing), and SARIMA models.

### ## Dataset

The dataset for the "Future Sales Prediction" project can be found on

Kaggle using the following link:

Future Sales Prediction Dataset.

**\*\*DATASET LINK:\*\***

<https://www.kaggle.com/datasets/chakradharmattapalli/future-sales-prediction>

The dataset contains the following columns:

**\*\*TV:\*\***

This column represents the amount of money spent on advertising through television. Television advertising is known for its wide reach and the potential to create awareness among a large and diverse audience. The values in this column indicate the advertising budget allocated to TV for different marketing campaigns or time periods.

**\*\*Radio:\*\***

This attribute reflects the advertising expenditure on radio campaigns. Radio advertising is often used to target specific demographics or regions, and its effectiveness can vary based on factors such as the choice of radio stations and time slots for airing advertisements.

#### **\*\*Newspaper:\*\***

This column measures the advertising spending on newspaper advertisements. Newspaper ads are typically employed for local or regional targeting and can come in various formats, sizes, and placements within the newspaper.

#### **\*\*Sales:\*\***

This is the dependent variable in the dataset, representing the actual sales figures achieved during or after each advertising campaign. The primary objective of analyzing this data is to understand how the budgets allocated to TV, Radio, and Newspaper advertising impact sales, allowing businesses to make informed marketing decisions

### **## Dependencies**

Before running the code, make sure you have the following dependencies installed:

- Python 3.x
- Pandas
- NumPy
- Seaborn
- Scikit-Learn
- Matplotlib
- Statsmodels
- XGBoost
- LightGBM
- TensorFlow (for the LSTM model)

You can install most of these dependencies using `pip`:

### **## How to Run the Code**

1. **\*\*Clone the Repository\*\***: Clone the repository to your local machine using Git. Replace `your-username` with your actual GitHub username.

```
```bash
```

```
git clone https://github.com/your-username/future-sales-prediction.git
...
```

2. **Navigate to the Project Directory**: Use the `cd` command to change your working directory to the project folder.

```
```bash
cd future-sales-prediction
...
```

3. **Prepare Your Sales Data**: Place your sales data in a CSV file named 'Sales.csv' in the project directory. Ensure that your data is structured with columns for 'TV,' 'Radio,' 'Newspaper,' and 'Sales.'

4. **Choose Execution Method**:

- **Using Jupyter Notebook**: If you prefer to run the code interactively and explore the results, you can use Jupyter Notebook. Run the following command:

```
```bash
jupyter notebook Sales_Prediction.ipynb
...
```

- **Using Python Script**: If you want to run the code as a standalone script, you can use the Python script. Execute the following command:

```
```bash
python Sales_Prediction.py
...
```

5. **Execute the Code**: Depending on your choice in the previous step, the code will either open in a Jupyter Notebook or run as a Python script. The code will load the data, preprocess it, and perform various analyses and predictions.

6. **View Results**: The code will display the results in the notebook or the terminal. You can analyze the model performance, view visualizations, and access the sales predictions.

That's it! You have successfully run the Future Sales Prediction code and obtained the results.

## ## Code Explanation

### **Data Loading and Preprocessing**

The code begins by loading the sales data from a CSV file and splitting it into features (X) and the target variable (y). It also performs data exploration and visualization.

### **\*\*Time Series Forecasting:\*\***

ARIMA, ETS, and SARIMA models are used to perform time series forecasting and generate forecasts for future sales.

### **\*\*Machine Learning Models:\*\***

Various machine learning models are trained and evaluated for sales prediction, including Linear Regression, Random Forest, Support Vector Machine, Neural Network (LSTM), XGBoost, and LightGBM.

### **\*\*Data Visualization:\*\***

The code includes various data visualization techniques such as histograms, scatter plots, box plots, density plots, and correlation matrices to understand the data and model performance.

### **\*\*Model Evaluation:\*\***

The code calculates and displays regression metrics, including Mean Squared Error (MSE), Mean Absolute Error (MAE), and R-squared (R2) scores for each machine learning model.

```
```bash
```

```
pip install pandas numpy seaborn scikit-learn matplotlib statsmodels xgboost lightgbm tensorflow
```