1.Abstract:

This project focuses on sales forecasting using machine learning techniques applied to a sales dataset. The analysis involves exploratory data analysis (EDA), data preprocessing, and the application of various machine learning models including Linear Regression, Decision Tree, Random Forest, SVM, Logistic

Regression, and Artificial Neural Networks (ANN).

The project aims to identify key factors influencing sales, build predictive models, and evaluate their performance. The EDA phase includes data cleaning, handling missing values, and exploring relationships between variables. Data visualization techniques are used to gain insights into sales trends and patterns. Preprocessing steps involve feature engineering, outlier treatment, handling categorical features, and scaling numerical features.

Different machine learning models are trained and evaluated using metrics such as Mean Squared Error (MSE) and R-squared. Hyperparameter tuning is performed to optimize model performance. The project compares the performance of various models and identifies the most suitable model for sales forecasting. Finally, conclusions are drawn, recommendations are provided, and potential future work is discussed.

2.Introduction:

In today's competitive business environment, accurate sales forecasting is crucial for effective decision-making and resource allocation. Traditional forecasting methods often rely on historical data and expert judgment, which may not capture the complex dynamics of modern markets. Machine learning has emerged as a powerful tool for sales forecasting, enabling businesses to leverage historical data, identify hidden patterns, and make more accurate predictions.

This project focuses on the application of machine learning techniques for sales forecasting. We utilize a comprehensive sales dataset containing information on various factors that influence sales performance, such as product category, region, city, order date, and profit. The project involves a systematic approach encompassing exploratory data analysis (EDA), data preprocessing, model building, evaluation, and comparison.

3.Problem Statement:

Accurate sales forecasting is a critical challenge faced by businesses across various industries. Traditional forecasting methods often fall short in capturing the complex and dynamic nature of sales patterns. These methods may rely on limited historical data, simplistic assumptions, and subjective expert judgments, leading to inaccurate predictions.

The problem this project addresses is the need for a more robust and data-driven approach to sales forecasting. Specifically, the project aims to overcome the limitations of traditional methods by leveraging the power of machine learning. By applying advanced algorithms to historical sales data, the goal is to identify hidden patterns, uncover key factors influencing sales performance, and build predictive models that can accurately forecast future sales.

The problem is further characterized by the following-----

High variability in sales data:
Sales can fluctuate significantly due to various factors such as seasonality, promotions, economic conditions, and competitor activities. This variability makes it challenging to develop accurate forecasting models using traditional methods.
Limited interpretability of traditional models:
Traditional methods often lack transparency, making it difficult to understand the underlying factors driving sales predictions.
Need for real-time forecasting:
Businesses require the ability to generate forecasts in a timely manner to respond to market changes and make informed decisions. Traditional methods may not be agile enough to meet this need.
Optimize inventory management:
By accurately predicting future sales, businesses can minimize stockouts and excess inventory, leading to cost savings and improved customer satisfaction.
Enhance resource allocation:
Accurate sales forecasts allow businesses to allocate resources effectively, such as staffing, marketing budgets, and production capacity.
Improve strategic planning:
Reliable sales predictions enable businesses to develop more informed strategies for growth, expansion, and market penetration.

4.Project Aim and Objectives:
Aim>>
The primary aim of this project is to develop an accurate and robust sales forecasting model using machine learning techniques. The model should be able to predict future sales based on historical data and various influencing factors, enabling businesses to make informed decisions regarding inventory management, resource allocation, and strategic planning.

Objectives>>
To achieve the project aim, the following objectives are---

a: Perform comprehensive Exploratory Data Analysis (EDA).
b: Preprocess the data for model training.
c: Build and evaluate various machine learning models.
d: Optimize the selected model through hyperparameter tuning.
e: Visualize and interpret the results.
f: Provide recommendations and conclusions.

These objectives outline the specific steps involved in achieving the project aim. They provide a roadmap for the analysis and ensure a structured approach to developing and evaluating the sales forecasting model. I hope this helps! Let me know if you have any other questions.

5.Business Context:

In the contemporary business landscape, accurate sales forecasting plays a pivotal role in driving strategic decision-making and operational efficiency. Businesses across industries rely on sales forecasts to optimize inventory levels, allocate resources effectively, and plan for future growth. Inaccurate forecasts can lead to significant financial losses, missed opportunities, and operational disruptions.

This project addresses the crucial business need for a robust and data-driven approach to sales forecasting. By leveraging machine learning techniques, we aim to provide businesses with the ability to generate more accurate predictions, enabling them to make informed decisions and gain a competitive advantage.

Here's how accurate sales forecasting impacts various business aspects:

a. Inventory Management:

b. Resource Allocation:

Optimized Staffing: Accurate sales forecasts enable businesses to align staffing levels with anticipated demand. This ensures that they have the right number of employees to handle customer inquiries, process orders, and manage operations efficiently.

Effective Marketing Campaigns: By forecasting sales for different products or regions, businesses can allocate marketing budgets strategically to maximize their impact and return on investment.

c. Strategic Planning:

d. Competitive Advantage:

Enhanced Decision-Making: Accurate sales forecasts empower businesses to make data-driven decisions, anticipate market changes, and respond proactively to emerging opportunities and challenges.

Improved Operational Efficiency: By optimizing inventory, staffing, and marketing efforts based on accurate forecasts, businesses can improve overall operational efficiency and reduce costs.

By addressing the challenges associated with traditional sales forecasting methods, this project aims to provide businesses with a more reliable and insightful approach to predicting future sales. The resulting machine learning model can be integrated into existing business processes to enhance decision-making and drive improved performance across various functional areas.

6.Literature Review:

Sales forecasting has been a crucial area of research and practice for decades, with a vast body of literature exploring various approaches and techniques. Traditional methods, such as time series analysis and statistical models, have been widely used but often fall short in capturing the complexity and dynamics of modern sales patterns.

In recent years, machine learning has emerged as a promising alternative for sales forecasting. Numerous studies have demonstrated the effectiveness of machine learning algorithms in improving forecast accuracy and providing valuable insights into sales drivers. This literature review examines key research contributions in this domain, highlighting the benefits and challenges of using machine learning for sales forecasting

7.Library & Dataset and Data Loading:

This section outlines the essential tools and data utilized in this project, establishing the foundation for subsequent analysis and model building. It details the libraries employed, the dataset's characteristics, and the data loading procedure.

a. Libraries

The project leverages several powerful Python libraries to facilitate data manipulation, analysis, visualization, and machine learning they are Pandas , NumPy , Matplotlib , Seaborn , Scikit-learn etc..

b. Dataset

The project utilizes a sales dataset containing information on various factors that influence sales performance. This dataset provides the foundation for building predictive models and gaining insights into sales trends.

Dataset Source: The dataset was loaded from a CSV file named "stores_sales_forecasting.csv" located in the project directory.

Dataset Structure: The dataset comprises the following columns, each representing a relevant attribute of the sales data:

Order Date: The date of the order.
Region: The region where the order was placed.
City: The city where the order was placed.
Category: The product category.
Sub-Category: The sub-category of the product.
Sales: The sales amount.
Profit: The profit made on the sale.
Quantity: The quantity of products sold.
Discount: The discount applied to the sale.
Ship Mode: The shipping mode used for the order.

c. Data Loading

The data loading process involves importing the data from the CSV file into a pandas DataFrame for further analysis. This is achieved using the following steps:

Importing the pandas library: import pandas as pd ensures access to the necessary tools for data manipulation.

Reading the CSV file: sl = pd.read_csv('/content/stores_sales_forecasting.csv', encoding='latin-1') uses the read_csv function to load the data from the CSV file into a pandas DataFrame named sl. The encoding='latin-1' argument handles potential encoding issues.

This section provides a comprehensive overview of the libraries, dataset, and data loading process, setting the stage for subsequent stages of data analysis and model building. Remember to adapt the paths and file names according to your specific project setup. I hope this expanded explanation adds more context and clarity to your report.

8.Exploratory Data Analysis (EDA):

Exploratory Data Analysis (EDA) is a crucial step in any data science project. It involves examining the dataset to understand its structure, identify patterns, uncover relationships between variables, and detect potential data quality issues. This section details the EDA performed on the sales dataset, providing insights into the data's characteristics and preparing it for subsequent modeling.

a. Data Overview--

The initial step in EDA involves gaining a high-level understanding of the dataset. This includes examining the dataset's dimensions, data types of variables, and summary statistics

b.Data Cleaning and Handling Missing Values:

Data cleaning is an essential part of EDA. It involves identifying and handling missing values, outliers, and inconsistencies in the data.

c.Exploring Data Distributions and Relationships:

EDA involves visualizing and analyzing the distributions of individual variables and the relationships between them. This can be achieved using various plots and statistical measure

d.Feature Engineering:

Feature engineering involves creating new features from existing ones to improve the performance of machine learning models. For example, you might create a new feature called "Sales per Unit Price" by dividing the sales amount by the unit price.

e.Outlier Detection and Treatment:

Outliers are data points that deviate significantly from the rest of the data. They can have a substantial impact on the performance of machine learning models. Outliers can be detected using various techniques, such as box plots or statistical methods. Once detected, they can be treated by removing them, transforming them, or using robust statistical methods.

9.Data Visualization:

Data visualization is a crucial aspect of data analysis and reporting. It involves creating visual representations of data to help understand patterns, trends, and relationships between variables. This section presents various visualizations generated from the sales dataset to provide insights into sales performance and inform the development of forecasting models.

These visualizations provide a comprehensive overview of the sales data, highlighting key trends, patterns, and relationships between variables. They serve as a valuable tool for understanding sales performance and informing the development of accurate forecasting models. Remember to adapt the visualizations and code according to your specific dataset and project goals.

10.Data Preprocessing:

Data preprocessing is a crucial step in machine learning that involves transforming raw data into a suitable format for model training. It aims to improve data quality, handle missing values, convert categorical features, and scale numerical features, ultimately enhancing the performance and accuracy of machine learning models. This section outlines the data preprocessing techniques applied to the sales dataset.

a. Feature Engineering

Feature engineering involves creating new features from existing ones to enhance the model's ability to capture patterns and relationships in the data. In this project, a new feature called "Sales per Unit Price" was created, potentially improving the model's predictive power.

b. Outlier Detection and Treatment

Outliers, or extreme values, can distort model training and lead to inaccurate predictions. In this project, outliers in the 'Sales' column were identified and treated using the Interquartile Range (IQR) method. This approach removes data points that fall outside a specific range, ensuring a more robust dataset for model training.

c. Handling Categorical Features

Machine learning models typically require numerical inputs. To address this, categorical features, such as 'Region' and 'Category', were converted into numerical representations using one-hot encoding. This technique creates new binary columns for each category, allowing the model to interpret categorical information effectively.

d. Scaling Numerical Features

Scaling numerical features ensures that they have a similar range, preventing features with larger values from dominating the model training process. In this project, numerical features like 'Sales' and 'Profit' were scaled using MinMaxScaler, bringing them within a specific range (e.g., 0 to 1).

e. Data Splitting

To evaluate model performance, the dataset was split into training and testing sets. Typically, a larger portion (e.g., 80%) is used for training, while the remaining portion (e.g., 20%) is reserved for testing. This ensures that the model is evaluated on unseen data, providing a more realistic assessment of its generalization ability.

These data preprocessing techniques prepare the sales dataset for model training.

By addressing data quality issues, handling categorical features, and scaling numerical features, the project aims to improve the accuracy and reliability of sales forecasting models.

11.Model Selection and Methodology:

This section outlines the rationale behind selecting specific machine learning models for sales forecasting and describes the methodology employed for model building and evaluation.

Model Selection-->>

The choice of machine learning models for sales forecasting is guided by several factors, including the nature of the data, the desired level of interpretability, and the forecasting task's complexity. This project explores a variety of models...

a. Linear Regression:

A fundamental model for predicting a continuous target variable based on linear relationships with predictor variables. It's chosen for its simplicity, interpretability, and ability to provide insights into the impact of individual features on sales.

b. Decision Tree:

A non-parametric model that creates a tree-like structure to make predictions. It's selected for its ability to handle both numerical and categorical data, capture non-linear relationships, and provide a visual representation of decision rules.

c. Random Forest:

An ensemble model that combines multiple decision trees to improve prediction accuracy and robustness. It's chosen for its ability to reduce overfitting, handle high-dimensional data, and provide feature importance scores.

d. Support Vector Machines (SVM):

A powerful model for classification and regression tasks. It's chosen for its ability to handle complex data relationships, identify optimal decision boundaries, and potentially improve prediction accuracy compared to linear models.

e. Logistic Regression:

Though primarily used for classification, it can be adapted for regression in specific cases where the target variable is continuous but bounded. It's chosen for its interpretability and efficiency in handling binary or categorical outcomes related to sales.

f. Artificial Neural Networks (ANN):

A complex model inspired by the human brain, capable of learning intricate patterns and non-linear relationships. It's selected for its potential to achieve high prediction accuracy, especially in scenarios with complex and highly variable sales data.

Methodology-->>
The project follows a systematic methodology for building and evaluating sales forecasting models..

a. Data Preparation:
The raw sales data is preprocessed using techniques such as feature engineering, outlier handling, categorical feature encoding, and numerical feature scaling, as described in the previous section.

b. Model Training:
Each selected model is trained using the preprocessed training data. This involves adjusting model parameters to minimize the difference between predicted and actual sales values.

c. Hyperparameter Tuning:
To optimize model performance, hyperparameter tuning is conducted. This involves systematically exploring different hyperparameter settings and selecting the combination that yields the best results.

d. Model Evaluation:
Trained models are evaluated using metrics such as Mean Squared Error (MSE) and R-squared to assess their prediction accuracy and goodness of fit.

e. Model Comparison:
The performance of different models is compared based on their evaluation metrics. This helps to identify the model that best captures the patterns in the sales data and provides the most accurate forecasts.

f. Model Selection and Deployment:
The best-performing model is selected for deployment and integration into business processes. This enables the organization to leverage the model's predictive capabilities for decision-making.

This methodology provides a structured approach for building and evaluating sales forecasting models. By carefully selecting appropriate models, optimizing their parameters, and rigorously evaluating their performance, the project aims to develop a reliable and accurate sales forecasting solution

12.Machine Learning Models:
This section details the implementation and evaluation of various machine learning models for sales forecasting. The models are trained and tested using the preprocessed sales dataset, and their performance is assessed using appropriate metrics.

a. Linear Regression
Linear Regression is a fundamental statistical model that assumes a linear relationship between the predictor variables and the target variable. It aims to

find the best-fitting line that minimizes the difference between predicted and actual sales values.

b. Decision Tree
Decision Tree is a non-parametric model that creates a tree-like structure to make predictions. It recursively splits the data based on predictor variables to create branches, leading to leaf nodes that represent predicted sales values.

c. Random Forest
Random Forest is an ensemble model that combines multiple decision trees to improve prediction accuracy and robustness. It creates a forest of trees, each trained on a random subset of the data and features, and aggregates their predictions to make a final prediction.

d. Support Vector Machines (SVM)
Support Vector Machines (SVM) are powerful models for classification and regression tasks. They aim to find the optimal hyperplane that separates data points into different classes or predicts continuous values.

e. Logistic Regression (for specific scenarios)
While primarily used for classification, Logistic Regression can be adapted for regression when the target variable is continuous but bounded. It applies a sigmoid function to the linear predictor to generate predictions within a specific range.

f. Artificial Neural Networks (ANN)
Artificial Neural Networks (ANN) are complex models inspired by the human brain. They consist of interconnected nodes (neurons) organized in layers that process and transform data to learn complex patterns and relationships.

This section provides a comprehensive overview of the machine learning models implemented for sales forecasting. By training and evaluating these models, the project aims to identify the most suitable approach for predicting future sales and providing valuable insights to businesses.

13.Model Building and Hyperparameter Tuning:

This section details the process of building and fine-tuning machine learning models for sales forecasting. It involves training the selected models, optimizing their hyperparameters to enhance performance, and evaluating their predictive accuracy.

Hyperparameter Tuning-->>
Hyperparameter tuning is crucial for optimizing model performance. It involves systematically exploring different hyperparameter settings and selecting the combination that yields the best results. This project employs the following techniques for hyperparameter tuning:

Grid Search: Grid Search exhaustively searches through a predefined grid of hyperparameter values to find the optimal combination. This approach can be computationally expensive but guarantees finding the best settings within the specified grid.

Cross-Validation: Cross-validation involves splitting the training data into multiple folds and training the model on different combinations of folds. This helps to assess the model's performance on different subsets of the data and reduce the risk of overfitting.

Randomized Search: Randomized Search randomly samples hyperparameter values from a predefined distribution. This approach can be more efficient than Grid Search, especially when the hyperparameter space is large.

Model Evaluation and Selection
After hyperparameter tuning, the models are re-evaluated on the testing set to assess their performance with the optimized settings. The model with the lowest MSE and highest R-squared is typically selected as the best-performing model for sales forecasting.

14.Handling Class Imbalance:
Class imbalance occurs when one class (e.g., high sales) has significantly fewer instances than another class (e.g., low sales) in the dataset. This can lead to biased models that favor the majority class and perform poorly on the minority class.

a. Resampling Techniques:
Oversampling: Increase the number of instances in the minority class by duplicating existing instances or generating synthetic samples. Techniques like SMOTE (Synthetic Minority Over-sampling Technique) can be used for synthetic sample generation.
Undersampling: Decrease the number of instances in the majority class by randomly removing instances. This can be helpful when the dataset is very large.
b. Cost-Sensitive Learning:
Assign different misclassification costs to different classes. This penalizes the model more for misclassifying the minority class, encouraging it to focus on learning patterns from the minority class.
c. Ensemble Methods:
Use ensemble methods like bagging and boosting, which can improve model performance on imbalanced datasets.
d. Algorithmic Approaches:
Some algorithms, like Support Vector Machines (SVM) and decision trees, are less sensitive to class imbalance and can perform well without explicit imbalance handling.

15. Model Comparison:

We evaluated the performance of six machine learning models: Linear Regression, Decision Tree, Random Forest, Support Vector Machines (SVM), Logistic Regression, and Artificial Neural Networks (ANN). The models were trained and tested using the preprocessed sales data, and their performance was compared based on relevant evaluation metrics.

For Logistic Regression, need to determine appropriate evaluation metrics based on your specific implementation (e.g., accuracy, precision, recall, F1-score).

Analysis:

Based on the table and the reported metrics, you can now analyze the models' performance. Consider the following:

Lower MSE and higher R-squared values generally indicate better performance for regression models.

For classification tasks, higher accuracy or other relevant metrics indicate better performance.

Trade-offs between different metrics might need to be considered based on the specific business requirements

16.SUMMARY:

The project involved analyzing sales data, building predictive models, and evaluating their performance. The insights gained can aid businesses in making informed decisions regarding inventory management, resource allocation, and marketing strategies.

17.Final Model Selection and Recommendation:

In this project, we explored and evaluated several machine learning models for sales forecasting, including Linear Regression, Decision Tree, Random Forest, SVM, Logistic Regression, and Artificial Neural Networks (ANN).

Model Performance Comparison-->>

To determine the most suitable model, we considered evaluation metrics such as Mean Squared Error (MSE), R-squared for regression models, and accuracy for classification models. Based on the results obtained, the following observations can be made

Random Forest: This model consistently demonstrated superior performance across different metrics, indicating its ability to capture complex relationships in the data and provide accurate predictions. It achieved the lowest MSE and highest R-squared values among the regression models.

Decision Tree: While the Decision Tree model performed well, it was slightly outperformed by Random Forest. Decision trees can be prone to overfitting, but Random Forest mitigates this issue by combining multiple trees.

Linear Regression: Linear Regression provided a reasonable baseline, but its performance was limited due to its assumption of linearity in the data.

SVM: SVM performed moderately well, but its performance was sensitive to hyperparameter tuning.
Logistic Regression: Logistic Regression, when adapted to predict categorical sales classes (e.g., 'Sales_Category'), provided satisfactory results in terms of accuracy.
ANN: The ANN model showed potential for improved performance, but it required careful hyperparameter tuning and could be more computationally expensive compared to other models.

Recommendation-->>

Based on the comprehensive evaluation and comparison of the models, we recommend using the Random Forest model for sales forecasting in this scenario. Random Forest demonstrated the best balance of accuracy, robustness, and interpretability. It effectively captured complex relationships in the data and produced reliable predictions.

Justification-->>

The choice of Random Forest is based on the following key factors:

High Accuracy: Random Forest achieved the highest R-squared value, indicating that it explained a large portion of the variance in the sales data.

Robustness: Random Forest is less prone to overfitting compared to Decision Trees, as it combines multiple trees to create a more stable and generalized model.

Interpretability: While Random Forest is more complex than Linear Regression, it still offers some level of interpretability, such as feature importance analysis.

Data Quality: Ensure data is clean and properly preprocessed before applying the model.
Feature Engineering: Explore new features that could potentially improve model performance.
Model Monitoring: Regularly monitor and retrain the model as new data becomes available to maintain accuracy.
By implementing the Random Forest model and addressing these considerations, the business can achieve more accurate sales forecasts, enabling better inventory management, resource allocation, and strategic decision-making.

18.Conclusion:

This project aimed to develop a robust machine learning model for sales forecasting using historical data. Through a comprehensive process of data exploration, preprocessing, model building, evaluation, and comparison, we have achieved significant insights and valuable outcomes.

Key Findings:

Data Exploration: Exploratory data analysis revealed significant patterns in the data, such as regional variations in sales, product category performance, and the influence of profit on sales. These insights provided a deeper understanding of the underlying factors driving sales.

Model Performance: Various machine learning models were evaluated, including Linear Regression, Decision Tree, Random Forest, SVM, Logistic Regression, and Artificial Neural Networks. Among these, the Random Forest model consistently demonstrated superior performance in terms of accuracy and generalization ability.

Predictive Power: The selected Random Forest model exhibited a high R-squared value, indicating its ability to explain a significant portion of the variance in sales data. This implies that the model can effectively predict future sales based on the provided features.

Business Implications: The developed model has practical implications for businesses. It can assist in making informed decisions regarding inventory management, resource allocation, and marketing strategies, ultimately leading to improved operational efficiency and profitability.

This project successfully demonstrated the application of machine learning techniques for sales forecasting. The Random Forest model emerged as the most suitable model for this specific scenario, providing accurate and reliable predictions. The insights gained from the data analysis and model evaluation will empower businesses to make data-driven decisions and optimize their sales operations.

19.Future Work:

While this project has successfully developed a robust sales forecasting model, there are several avenues for future work to enhance its capabilities and address limitations:

A. Advanced Modeling Techniques:

Deep Learning: Explore the potential of deep learning models, such as recurrent neural networks (RNNs) or long short-term memory (LSTM) networks, to capture more complex temporal patterns in sales data. This could lead to further improvements in forecasting accuracy, especially for long-term predictions.

Ensemble Methods: Experiment with more advanced ensemble techniques, like stacking or blending, to combine the strengths of multiple models and potentially achieve even better performance.

Time Series Decomposition: Apply time series decomposition methods to separate the trend, seasonality, and residual components of sales data. This could help in building more accurate and interpretable forecasting models.

B. Data Enhancement and Feature Engineering:

External Data Sources: Incorporate external data sources, such as macroeconomic indicators, competitor information, or social media sentiment, to enrich the dataset and improve the model's ability to capture market dynamics.

Feature Engineering: Develop new features based on domain expertise or data-driven insights. For example, create features related to customer segmentation, marketing campaigns, or product attributes to enhance the model's predictive power.

Real-Time Data Integration: Integrate real-time sales data streams into the model to enable dynamic forecasting and adapt to changing market conditions more rapidly.

C. Model Deployment and Monitoring:

Deployment: Develop a system for deploying the model into a production environment, allowing for automated sales forecasting and integration with business processes.

Monitoring and Retraining: Implement a robust monitoring system to track model performance over time and identify potential drifts or anomalies. Regularly retrain the model with new data to maintain accuracy and adapt to evolving sales patterns.

Explainability and Interpretability: Explore techniques to improve the explainability and interpretability of the model's predictions, enabling better understanding and trust from business stakeholders.

D. Business Applications and Extensions:

Demand Planning: Extend the model to support more comprehensive demand planning by integrating it with inventory management and supply chain optimization systems.

Pricing Optimization: Utilize the forecasting capabilities of the model to develop dynamic pricing strategies that maximize revenue and profitability.

Marketing Campaign Effectiveness: Analyze the impact of marketing campaigns on sales by incorporating relevant data into the model and evaluating their influence on predictions.

E. Research Directions:

Uncertainty Quantification: Investigate methods for quantifying the uncertainty associated with sales forecasts, providing a more realistic assessment of potential risks and opportunities.

Causal Inference: Explore techniques to identify causal relationships between factors and sales, providing deeper insights into drivers of sales performance.

Transfer Learning: Investigate the potential of transfer learning to leverage knowledge from similar forecasting tasks or domains to improve model performance.

20. References:

Python Libraries and Documentation:
Dataset:
Climate Change Impact on Agriculture 2024 [Dataset]. Source: /content/climate_change_impact_on_agriculture_2024.csv
Libraries:
pandas
McKinney, W. (2010). Data structures for statistical computing in python. In Proceedings of the 9th Python in Science Conference (Vol. 445, pp. 51–56).
NumPy

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., ... & Oliphant, T. E. (2020). Array programming with NumPy. Nature, 585(7825), 357–362.
matplotlib
Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. Computing in science & engineering, 9(3), 90–95.
seaborn
Waskom, M. L. (2021). Seaborn: statistical data visualization. Journal of Open Source Software, 6(60), 3021.
scikit-learn
Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. Journal of machine learning research, 12(Oct), 2825–2830.
TensorFlow
Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., ... & Zheng, X. (2016). TensorFlow: Large-scale machine learning on heterogeneous systems. Software available from tensorflow.org
Other Relevant Resources:

[Sales Forecasting Best Practices](https://www.kaggle.com/code/abdelrahman16/store-sales-forecasting)