Pakistan Food Price Analysis

SRS Document

Table of content

Introduction:

Introduction:

The "Pakistan Food Price Analysis" project aims to conduct a comprehensive analysis of food prices in Pakistan using data mining techniques and algorithms. The project focuses on extracting valuable insights, identifying trends, and exploring relationships between different food and non-food items. By analyzing the "wfp dataset" sourced from the World Food Programme Price Database, we seek to understand the general trend in food prices over time, analyze significant relationships between various food commodities, assess the impact of increased oil prices on other food items, and compare price data across different provinces.

Pakistan, being a major producer of wheat and other food commodities, faces challenges of food insecurity despite its surplus production. With a significant percentage of the population experiencing food insecurity, it is crucial to gain deeper insights into the factors influencing food prices and their implications. This project aims to contribute to the understanding of the complex dynamics that shape food prices in Pakistan and provide valuable information for researchers and policymakers working towards addressing food security issues.

By employing data mining methodologies, we will process and analyze the "wfp dataset," which includes information on various food items, non-food items, and economic indicators. The project will leverage advanced algorithms such as time series analysis, association rule mining, and regression analysis to uncover patterns, correlations, and impacts within the dataset. Visualization techniques will be employed to present the analysis results in a clear and understandable manner.

The findings from this project have the potential to inform policy decisions, interventions, and strategies aimed at improving food security and ensuring better access to affordable and nutritious food for the Pakistani population. Additionally, researchers and analysts can benefit from the insights derived from the analysis to further explore the complexities of food price dynamics and contribute to the broader field of food security and agricultural economics.

Overall, the "Pakistan Food Price Analysis" project represents a significant endeavor to leverage data mining techniques to gain valuable insights into food price trends, relationships, and impacts in Pakistan. The outcomes of this analysis have the potential to contribute to evidence-based decision-making, facilitate targeted interventions, and foster a better understanding of the factors affecting food security in the country.

Scope:

The scope of the "Pakistan Food Price Analysis" project is to conduct a comprehensive analysis of food prices in Pakistan using data mining techniques and algorithms. The project aims to extract insights, identify trends, and explore relationships between different food and non-food items. The analysis will be performed on the "wfp dataset" sourced from the World Food Programme Price Database. The focus will be on understanding the general trend in food prices over time, analyzing the significant relationships between different food prices, exploring the impact of increased oil prices on other food items, and comparing data across different provinces to identify general trends.

Tools:

- Python (programing language)
- Anaconda
- Numpy
- Jupiter Lab.

Data source:

World Food Programme

Data file: https://data.humdata.org/dataset/wfp-food-prices-for-pakistan

References:

https://learn.microsoft.com/en-us/analysis-services/data-mining/data-mining-algorithms-analysis-services-data-mining?view=asallproducts-allversions

https://www.wfp.org/countries/pakistan

Algorithms:

We aim to implement the latest AI algorithms to achieve the highest accuracy and precision for the "Pakistan Food Price Analysis" project. We will explore renowned AI algorithms such as:

- Naive Bayes
- Apriori
- Association Rule Mining
- Decision Trees
- Support Vector Machines
- Logistic Regression
- C4.5
- CART
- CNN
- RSNET

and others. As a team, we will carefully consider the data needs and requirements to select the most suitable algorithms, ensuring valid reasons for their choice.

By leveraging AI algorithms, we can effectively analyze the "wfp dataset" and uncover valuable insights related to food prices in Pakistan. Each algorithm offers unique capabilities and strengths that can be advantageous depending on the specific analysis objectives. For instance, Naive Bayes is well-suited for probabilistic classification tasks, while Decision Trees excel in visualizing hierarchical relationships. Support Vector Machines are effective in handling high-dimensional data and logistic regression is useful for predicting categorical outcomes.

Our selection of algorithms will be driven by factors such as the type of data, the desired outcome (e.g., trend analysis, relationship identification), and the complexity of patterns within the dataset. We will consider the algorithm's ability to handle large volumes of data, its interpretability, computational efficiency, and its track record in similar data mining projects.

By employing a diverse range of AI algorithms, we aim to maximize the accuracy, precision, and comprehensiveness of our analysis. This approach will enable us to gain deep insights into food price

dynamics, identify significant relationships, and make informed recommendations to address food insecurity and improve accessibility to affordable and nutritious food in Pakistan.

Requirement

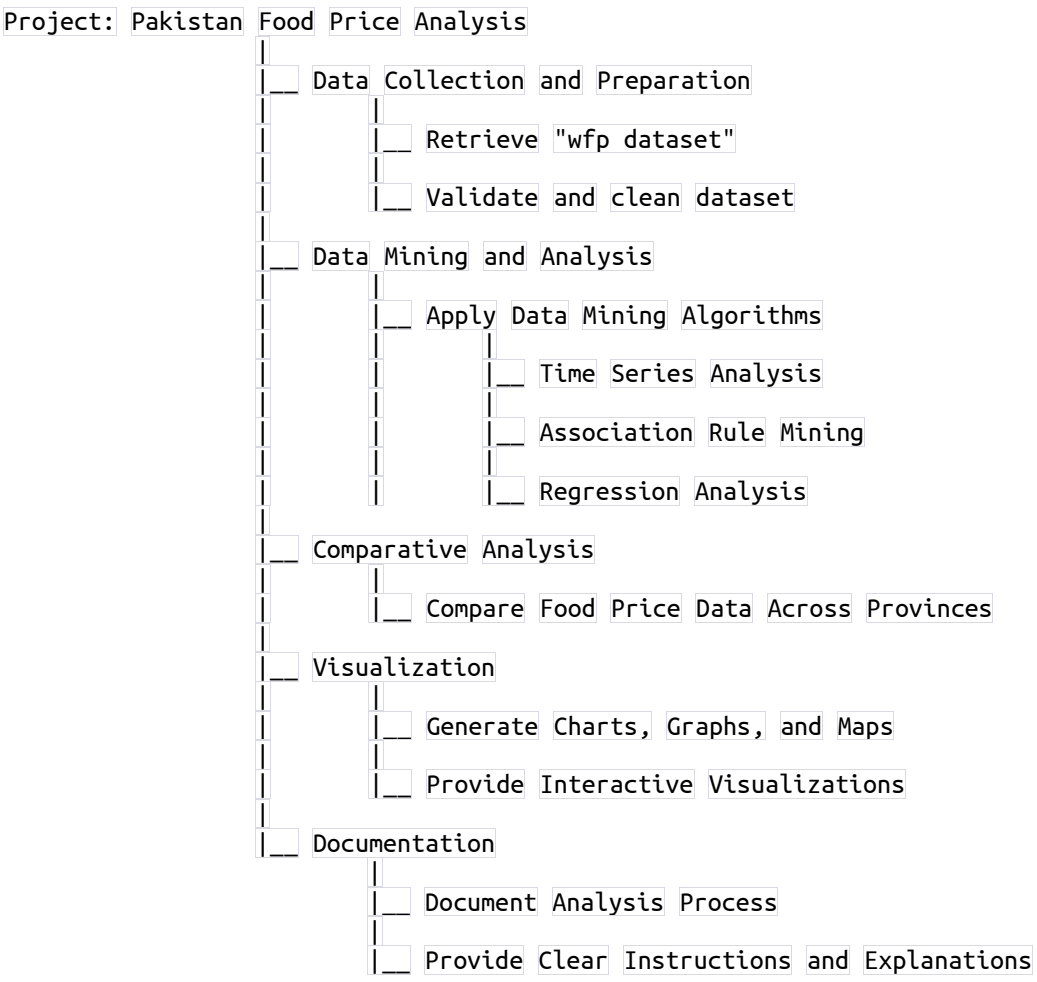
Functional Requirements:

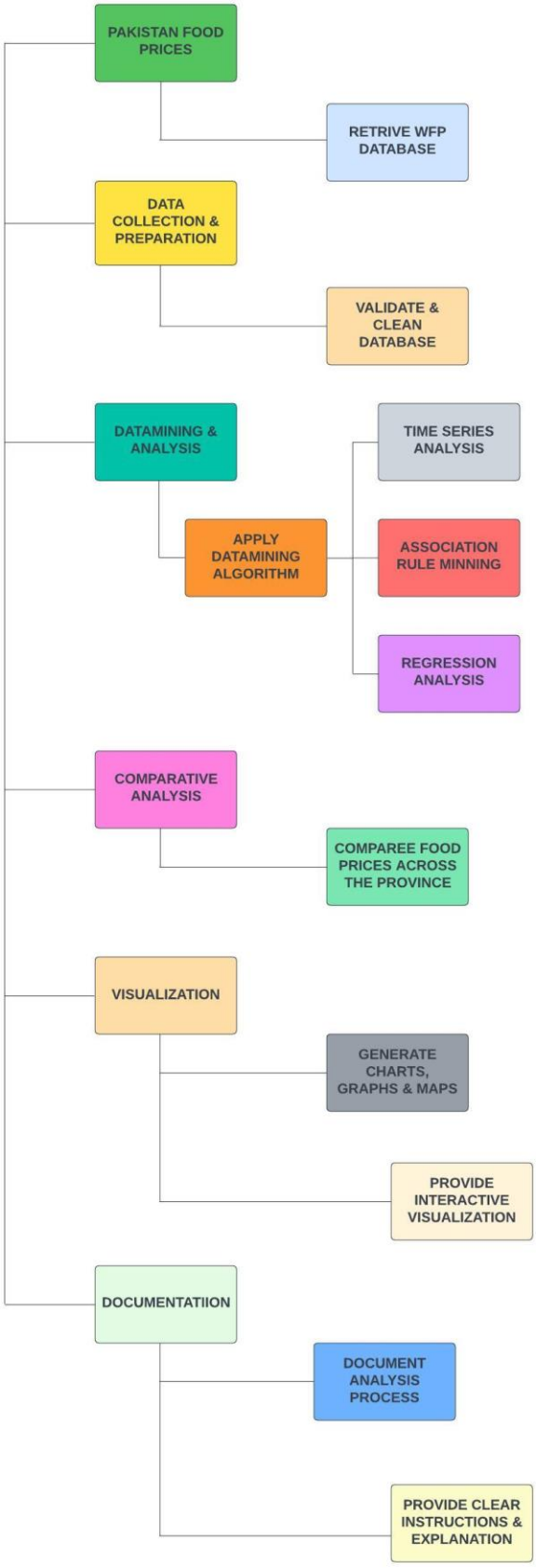
1. Trend Analysis:
    - The system should process the 'wfp dataset' to analyze the general trend in food prices over time in Pakistan.
    - It should identify and visualize the patterns, fluctuations, and overall direction of food prices for various commodities such as maize, rice, wheat, beans, fish, and sugar.
    - The analysis should consider both short-term and long-term trends to provide a comprehensive understanding of price movements.
2. Relationship Analysis:
    - The system should analyze the significant relationships between different food prices in Pakistan.
    - It should identify correlations, dependencies, and associations between food commodities to uncover patterns and interdependencies.
    - The analysis should help in understanding how changes in the price of one food item may impact the prices of other items.
3. Impact of Oil Price Increase:
    - The system should examine and analyze the impact of increased oil prices on the prices of other food items in Pakistan.
    - It should identify any direct or indirect relationships between oil prices and food prices.
    - The analysis should quantify the influence of oil price fluctuations on food prices and provide insights into the magnitude of the impact.
4. Provincial Comparison:
    - The system should compare food price data across different provinces of Pakistan over time.
    - It should analyze and visualize the general trends in food prices for each province.
    - The analysis should highlight any variations, similarities, or disparities in price movements among provinces, allowing for regional comparisons.

Non-functional Requirements:

1. Usability:
    - The analysis process and its results should be presented in a clear and understandable manner, even to non-technical stakeholders. The visualizations and reports should be user-friendly and intuitive.
    - The analysis should be easily replicable and modifiable to allow for future updates or modifications as needed.
2. Performance:
    - The data mining and analysis processes should be optimized for efficient execution, considering the large volume of data. The analysis should be performed within reasonable time limits to provide timely insights.
    - The system should be able to handle large datasets and complex computations without significant performance degradation.
3. Accuracy and Reliability:
    - The analysis should be accurate and reliable, ensuring that the results are consistent and meaningful.
    - The data preprocessing and cleaning steps should be thorough to minimize errors and inconsistencies in the analyzed data.
4. Security:
    - The data used for analysis should be handled securely to maintain confidentiality and protect against unauthorized access or data breaches.
    - Any sensitive information or personal data should be properly anonymized or masked.
5. Scalability:
    - The analysis framework should be scalable to accommodate potential future increases in data volume and complexity.
    - It should be able to handle additional data sources or variables if needed.
    - Documentation:
    - Comprehensive documentation should be provided, outlining the steps taken during the analysis process, including data preprocessing, algorithms used, and visualization techniques employed.
    - The documentation should be clear, well-structured, and accessible for future reference and replication.

Hierarchical diagram:




PAKISTAN FOOD PRICES
RETRIVE WFP DATABASE
DATA COLLECTION & PREPARATION
VALIDATE & CLEAN DATABASE
DATAMINING & ANALYSIS
TIME SERIES ANALYSIS
APPLY DATAMINING ALGORITHM
ASSOCIATION RULE MINNING
REGRESSION ANALYSIS
COMPARATIVE ANALYSIS
COMPAREE FOOD PRICES ACROSS THE PROVINCE
VISUALIZATION
GENERATE CHARTS, GRAPHS & MAPS
PROVIDE INTERACTIVE VISUALIZATION
DOCUMENTATIION
DOCUMENT ANALYSIS PROCESS
PROVIDE CLEAR INSTRUCTIONS & EXPLANATION

Usage Scenario:

Use Case	Use Case ID	Actions	Description
Retrieve and Validate Dataset	UC1	1. Access the World Food Programme Price Database. 2. Retrieve the "wfp dataset" for food prices in Pakistan. 3. Validate the dataset for accuracy and completeness. 4. Flag any missing values, outliers, or inconsistencies.	This scenario involves accessing the dataset from the World Food Programme Price Database, ensuring its accuracy and completeness by performing validation checks.
Perform Time Series Analysis	UC2	1. Select the time series analysis algorithm. 2. Configure the algorithm parameters (time intervals, smoothing techniques). 3. Apply the time series analysis algorithm to the food price data. 4. Identify general trends, seasonal variations, and patterns. 5. Generate visualizations to illustrate the findings.	This scenario focuses on analyzing the food price data using time series analysis techniques to identify trends, seasonal patterns, and visualizing the results.
Conduct Association Rule Mining	UC3	1. Choose the association rule mining algorithm. 2. Set algorithm parameters (minimum support, confidence thresholds). 3. Apply the association rule mining algorithm to identify relationships between food items. 4. Generate frequent itemsets and association rules. 5.	This scenario involves applying association rule mining techniques to discover relationships between different food items and presenting the findings using visualizations.

		Present the associations through visualizations.	
Analyze Impact of Increased Oil Prices	UC4	1. Select the regression analysis algorithm. 2. Define relevant variables and factors for the regression model. 3. Perform regression analysis on the food price and oil price data. 4. Quantify relationships between oil prices and other food item prices. 5. Generate visualizations to illustrate the findings.	This scenario focuses on analyzing the impact of increased oil prices on food prices using regression analysis techniques, quantifying relationships, and visualizing the results.

Adopted Methodology:

Data Mining Methodology:

1. Problem Definition: Clearly define the problem statement and objectives of your data mining project. Identify the specific questions you want to answer or the patterns and insights you want to discover from the food price dataset.

2. Data Collection: Gather the relevant data for analysis. This may involve acquiring food price data from the World Food Programme Price Database or other reliable sources. Ensure that the data is comprehensive, accurate, and representative of the desired time period and regions.

3. Data Understanding: Explore and familiarize yourself with the dataset. Understand the data's structure, variables, and quality. Identify any missing values, outliers, or data quality issues that need to be addressed.

4. Data Preprocessing: Cleanse and preprocess the data to ensure its quality and suitability for analysis. This may involve tasks such as handling missing values, dealing with outliers, removing duplicates, normalizing or scaling variables, and performing feature selection or engineering.

5. Data Mining Techniques: Apply appropriate data mining techniques and algorithms to analyze the dataset and extract insights. For example, you can use association rule mining to identify relationships between different food items, regression analysis to analyze the impact of increased oil prices on food prices, or time series analysis to discover trends over time.

6. Model Evaluation: Evaluate the performance and accuracy of the data mining models or algorithms. This may involve techniques such as cross-validation, model comparison, or hypothesis testing. Ensure that the chosen models effectively capture the patterns and relationships in the data.

7. Interpretation and Visualization: Interpret and communicate the results of your data mining analysis. Use appropriate data visualization techniques to present the findings in a clear and understandable manner. This can include charts, graphs, heatmaps, or maps to showcase the trends, relationships, and patterns discovered in the food price data.

8. Validation and Deployment: Validate the results and insights obtained from the data mining process. Ensure that the findings align with domain knowledge and make logical sense. Once validated, deploy the data mining results and insights for decision-making or further analysis.

9. Documentation and Reporting: Document all steps performed in the data mining process, including data preprocessing, algorithms used, parameters chosen, and results obtained. Prepare a comprehensive report summarizing the findings, insights, and recommendations derived from the data mining analysis.

Reasoning:

The Data Mining Methodology was chosen for the "Pakistan Food Price Analysis" project due to its suitability for data analytics and insights extraction. Unlike conventional system development methodologies, this project focuses primarily on data analysis rather than system building. The methodology offers flexibility, emphasizes analysis steps such as data understanding, preprocessing, mining techniques, evaluation, interpretation, visualization, validation, and documentation. Its iterative and agile approach allows for continuous refinement, exploration of algorithms, and optimization of analysis results. The chosen methodology aligns with the project's goal of delivering meaningful insights and recommendations through documentation and reporting, making it well-suited for this data analytics project.

Work Plan:

We have chosen to utilize Asana, Inc. as our project planning and tracking software for the "Pakistan Food Price Analysis" project. Asana provides a comprehensive set of features that are well-suited for data mining projects, specifically in terms of task management and visualization.

The key reasons for selecting Asana are as follows:

1. Task Organization: Asana allows us to create tasks and organize them into sections such as "Done," "Doing," and "To Do." This feature enables us to easily track the progress of tasks and maintain an organized workflow.

2. Milestone Tracking: Asana enables us to set milestones within tasks, providing a clear timeline and progress overview for the project. This feature allows us to track major project milestones and ensure timely completion.

3. Subtask Management: Asana facilitates the creation of subtasks, allowing for more detailed breakdown and management of complex tasks. This functionality helps us to efficiently manage and track various sub-components of the project.

4. Assignment and Collaboration: Asana allows us to assign tasks and subtasks to team members, ensuring clear responsibility allocation. It enables collaboration by providing a platform for team members to communicate, share updates, and collaborate on specific tasks or subtasks.

5. Priority and Due Dates: Asana allows us to assign priority levels and set due dates for tasks, subtasks, and milestones. This feature helps us prioritize tasks based on their importance and urgency, ensuring efficient task management.

6. Real-Time Updates: Asana provides real-time updates on task progress and status, enabling team members to stay informed about the project's current state. This feature fosters transparency and facilitates effective communication within the team.

In comparison to other software options like Jira, Asana offers a more user-friendly interface and a broader range of project management features specifically designed for task organization and

visualization. Additionally, Asana is more accessible to students as it offers affordable pricing plans and provides a free version with limited functionality.

Overall, Asana's capabilities align well with the needs of our data mining project, making it the ideal choice for project planning and tracking.