Federal Urdu University of Arts, Sciences & Technology

NAME: ABDUL SAMI KHAN
: SHAHEER AHMED

SECTION: A

SEAT NO: 19122107
19122089

PROJECT NAME: PAKISTAN FOOD PRICE

COURSE NAME: DATA MINING

SUBMITTED: MISS UZMA FAZAL

OBJECTIVE:

The primary report is to evaluate and understand the dynamics of food prices across various commodities, providing a holistic view of the market and Assess the impact of economic policies on the pricing of diverse food commodities, guiding policymakers in effective decision-making for economic stability and also Analyze the affordability of a range of essential food commodities, aiding in addressing cost-of-living concerns and ensuring accessibility to diverse and nutritious food items. The report aims to predict the commodities of food price with the association rule mining model and random forest model. The major column in the dataset are as follows:

*)Date
*)Province
*)City name
*)City market
*)Category
*)Commodity
*)Unit
*)Price type
*)Retail
*)Price

INTRODUCTION AND BACKGROUND OF THE PROBLEM:

The food price dataset is a comprehensive collection of information that captures the prices of various food items across different regions, markets, and time periods. This dataset is a valuable resource for researchers, policymakers, economists, and other stakeholders interested in understanding and analyzing trends, patterns, and factors influencing food prices. Fluctuations in food prices have significant implications for individuals, communities, and economies. The volatility in food prices can be

influenced by a multitude of factors, including weather conditions, agricultural practices, geopolitical events, market dynamics, and supply chain disruptions. Understanding and mitigating the impact of these factors is crucial for ensuring food security, economic stability, and the well-being of populations.

DATA COLLECTION:

The data has been taken to analyze the rate of commodities from
https://opendata.com.pk/
https://opendata.com.pk/dataset?q=food+price

PRE-PROCESSING:

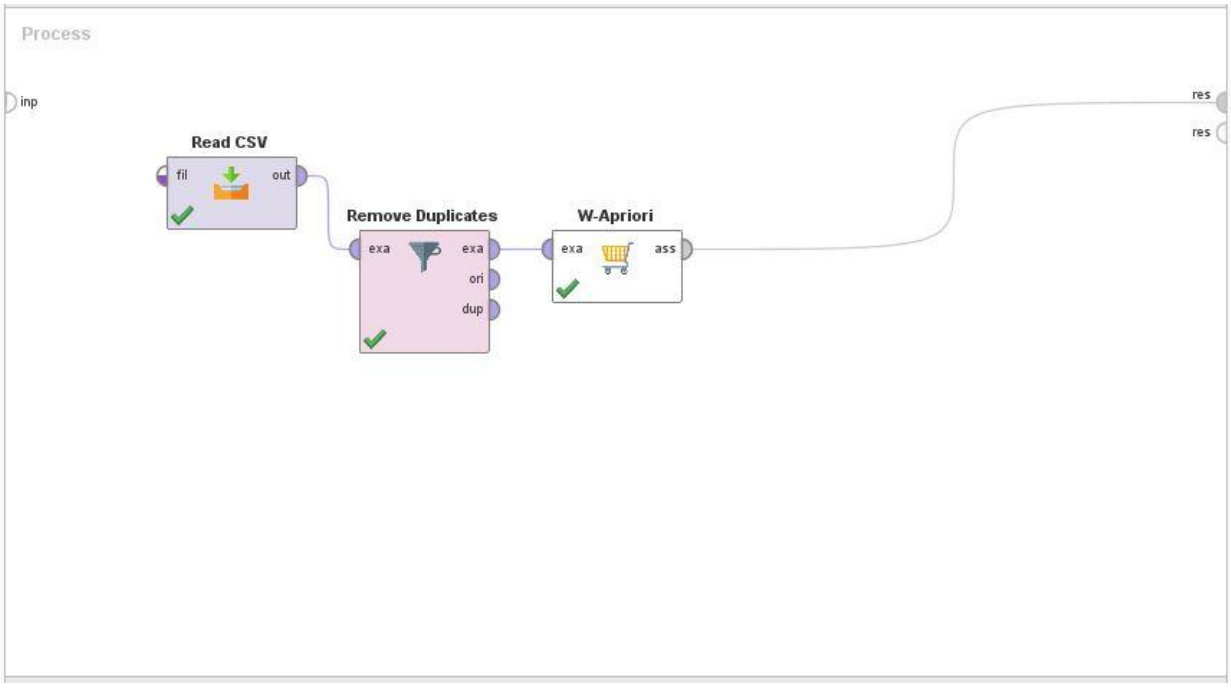
Before applying any model on the dataset first it undergoes on prep processing. It handle these steps:

*)Data Cleaning
*)Handle Missing Values
*)Remove Duplicate Values

MODELING:

*)ASSOCIATION RULE MINING:

Association rule mining is a technique used to discover interesting relationships, patterns, and associations within a dataset. In the context of a food price dataset, association rule mining can help identify significant associations between different food commodities or patterns in their pricing behavior. The Apriori algorithm is a popular algorithm for association rule mining



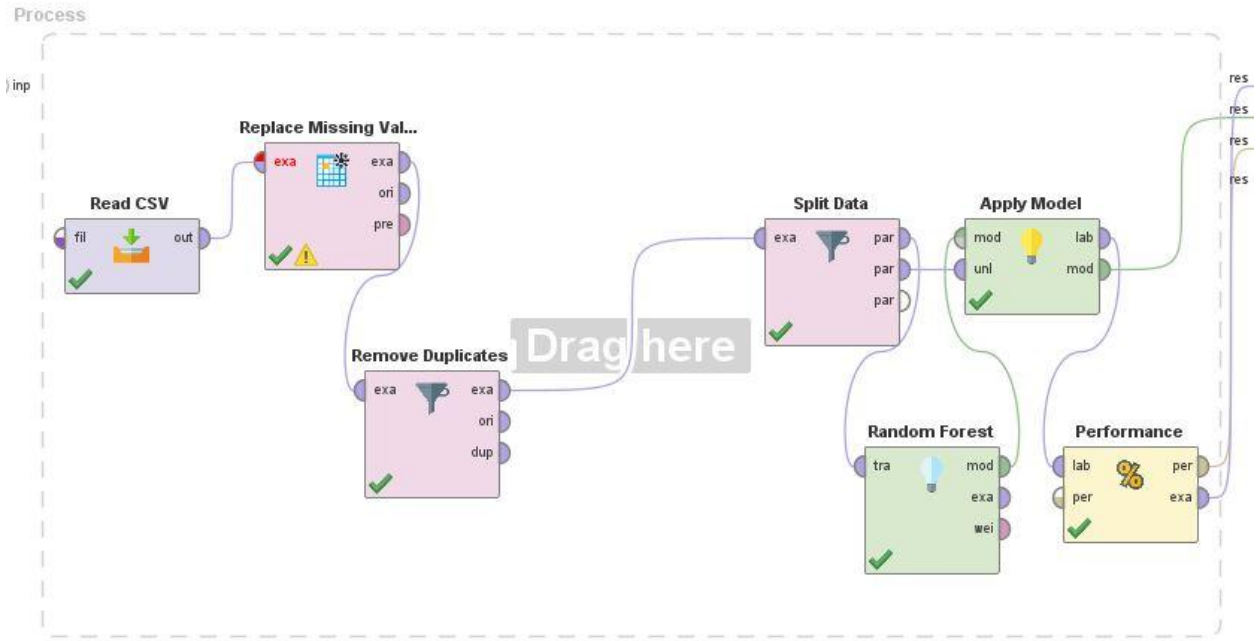
W-Apriori

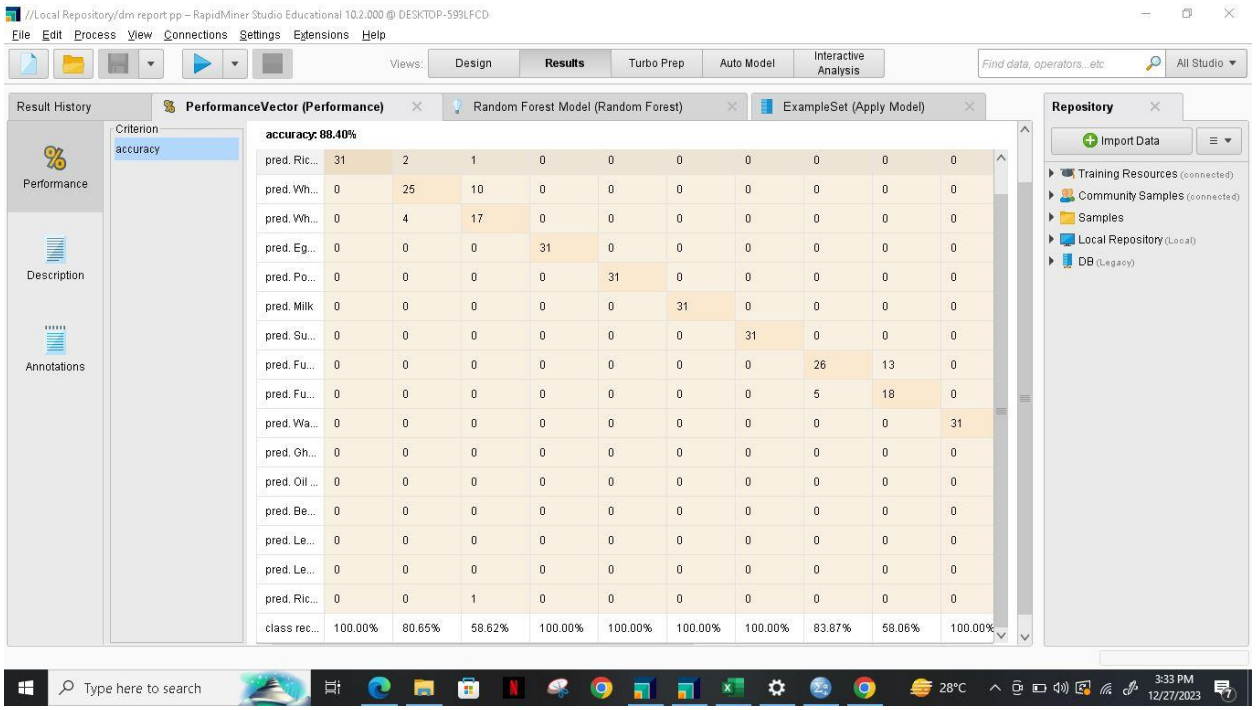
```
Apriori
=======

Minimum support: 0.2 (16 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 16

Generated sets of large itemsets:

Size of set of large itemsets L(1): 12

Size of set of large itemsets L(2): 10

Size of set of large itemsets L(3): 3

Best rules found:

 1. City Name=Quetta 16 ==> Provinces name=BALOCHISTAN 16    conf:(1)
 2. Provinces name=BALOCHISTAN 16 ==> City Name=Quetta 16    conf:(1)
 3. City market=Quetta 16 ==> Provinces name=BALOCHISTAN 16    conf:(1)
 4. Provinces name=BALOCHISTAN 16 ==> City market=Quetta 16    conf:(1)
 5. City Name=Peshawar 16 ==> Provinces name=KHYBER PAKHTUNKHWA 16    conf:(1)
 6. Provinces name=KHYBER PAKHTUNKHWA 16 ==> City Name=Peshawar 16    conf:(1)
 7. City market=Peshawar 16 ==> Provinces name=KHYBER PAKHTUNKHWA 16    conf:(1)
 8. Provinces name=KHYBER PAKHTUNKHWA 16 ==> City market=Peshawar 16    conf:(1)
 9. City Name=Karachi 16 ==> Provinces name=SINDH 16    conf:(1)
10. Provinces name=SINDH 16 ==> City Name=Karachi 16    conf:(1)
```

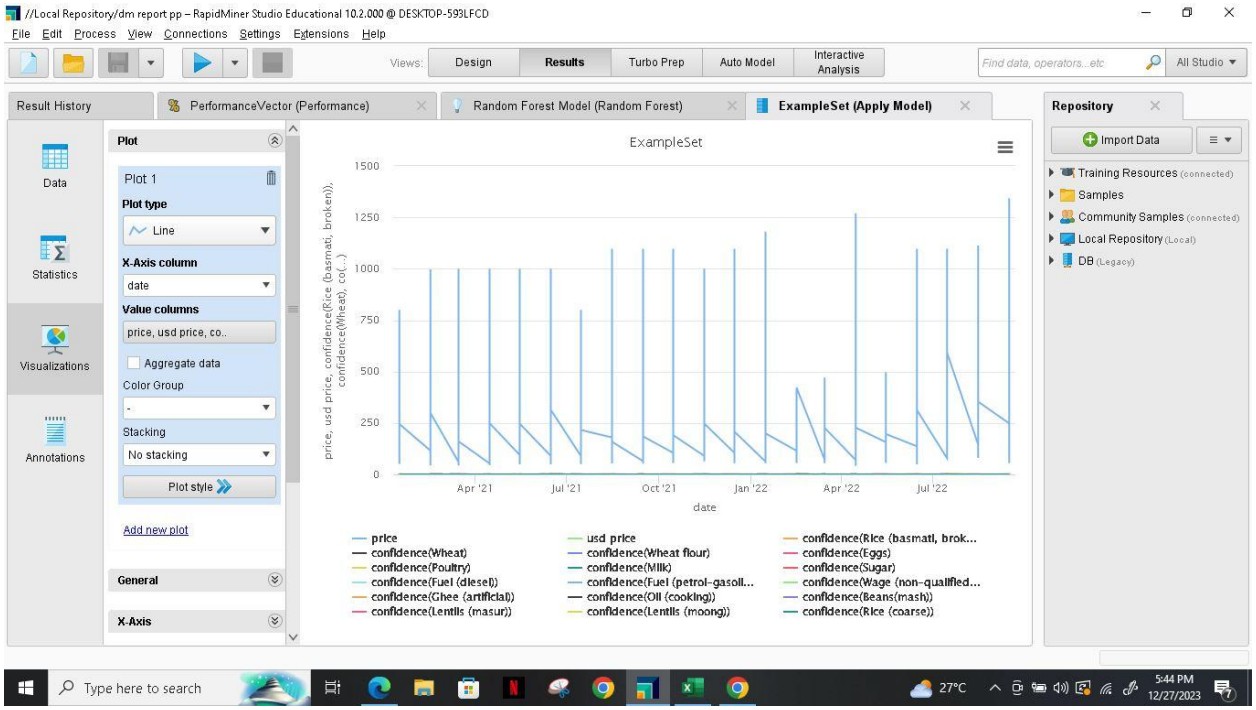
*)RANDOM FOREST:

Random forest is used for classification and regression task due to its high accuracy, feature importance and scalability, random forest reduce

overfitting.Using Random Forest in data mining with a food price dataset, especially when considering commodity labels, involves predicting or understanding price patterns based on various features and labels associated with commodities.




//Local Repository/dm report pp – RapidMiner Studio Educational 10.2.000 @ DESKTOP-593LFCD
File Edit Process View Connections Settings Extensions Help
Views: Design Results Turbo Prep Auto Model Interactive Analysis
Find data, operators...etc
All Studio
Result History
PerformanceVector (Performance)
Random Forest Model (Random Forest)
ExampleSet (Apply Model)
Repository
Import Data
Training Resources (connected)
Community Samples (connected)
Samples
Local Repository (Local)
DB (Legacy)
Performance
Description
Annotations
Criterion
accuracy
accuracy: 88.40%
pred. Ric... 31 2 1 0 0 0 0 0 0 0
pred. Wh... 0 25 10 0 0 0 0 0 0 0
pred. Wh... 0 4 17 0 0 0 0 0 0 0
pred. Eg... 0 0 0 31 0 0 0 0 0 0
pred. Po... 0 0 0 0 31 0 0 0 0 0
pred. Milk 0 0 0 0 0 31 0 0 0 0
pred. Su... 0 0 0 0 0 0 31 0 0 0
pred. Fu... 0 0 0 0 0 0 0 26 13 0
pred. Fu... 0 0 0 0 0 0 0 5 18 0
pred. Wa... 0 0 0 0 0 0 0 0 0 31
pred. Gh... 0 0 0 0 0 0 0 0 0 0
pred. Oil ... 0 0 0 0 0 0 0 0 0 0
pred. Be... 0 0 0 0 0 0 0 0 0 0
pred. Le... 0 0 0 0 0 0 0 0 0 0
pred. Le... 0 0 0 0 0 0 0 0 0 0
pred. Ric... 0 0 1 0 0 0 0 0 0 0
class rec... 100.00% 80.65% 58.62% 100.00% 100.00% 100.00% 100.00% 83.87% 58.06% 100.00%
Type here to search
28°C
3:33 PM
12/27/2023


//Local Repository/dm report pp – RapidMiner Studio Educational 10.2.000 @ DESKTOP-593LFCD
File Edit Process View Connections Settings Extensions Help
Views: Design Results Turbo Prep Auto Model Interactive Analysis
Find data, operators...etc
All Studio
Result History
PerformanceVector (Performance)
Random Forest Model (Random Forest)
ExampleSet (Apply Model)
Repository
Import Data
Training Resources (connected)
Samples
Community Samples (connected)
Local Repository (Local)
DB (Legacy)
Data
Statistics
Visualizations
Annotations
Plot
Plot 1
Plot type
Line
X-Axis column
date
Value columns
price, usd price, co..
Aggregate data
Color Group
-
Stacking
No stacking
Plot style
Add new plot
General
X-Axis
ExampleSet
1500
1250
1000
750
500
250
0
price, usd price, confidence(Rice (basmati, broken)), confidence(Wheat), co(...)
Apr '21
Jul '21
Oct '21
Jan '22
Apr '22
Jul '22
date
price
usd price
confidence(Rice (basmati, brok...
confidence(Wheat)
confidence(Wheat flour)
confidence(Eggs)
confidence(Poultry)
confidence(Milk)
confidence(Sugar)
confidence(Fuel (diesel))
confidence(Fuel (petrol-gasoli...
confidence(Wage (non-qualified...
confidence(Ghee (artificial))
confidence(Oil (cooking))
confidence(Beans(mash))
confidence(Lentils (masur))
confidence(Lentils (moong))
confidence(Rice (coarse))
Type here to search
27°C
5:44 PM
12/27/2023

RESULT:

The result of association rule mining gives the result of rules of commodities where the market based and random forest predict the prices of commodities with good accuracy.

CONCLUSION:

Summarize the overall findings and their significance in the context of the food industry. Emphasize how the analysis contributes to informed decision-making, potential improvements in operational efficiency, and strategies for navigating the complexities of the food market. Conclude with a call to action based on the recommendations provided.