PHASE: 2

MARKET BASKET INSIGHT

TEAM MEMBERS

310821106048: B. KAWIYA

310821106020: L. DHIVYA SHREE

310821106032: R. ISWARYA

310821106037: R. JEEVITHA

INTRODUCTION:

In data-driven decision-making, the key to the success of any project is the choice of the appropriate machine learning model, careful data processing and the quality of the data set. This project begins a journey through the complex landscape of model selection, data preparation and deep data exploration. This effort is based on the recognition that a well-chosen model can reveal hidden insights, predict future outcomes, and automate complex tasks. Equally important is data pre-processing, the process of transforming raw data into a refined, reliable and informative format. At the heart of it all, however, is the dataset Itself, a treasure trove of information waiting to be unlocked. Each aspect—model selection, Preprocessing, and data exploration—contributes significantly to the project's overall goal of using data to make informed decisions and gain valuable insights. The success of this project depends on the synergy of these components, making it an exciting exploration of the art and science of data analysis.

DATA PROCESSING:

In the initial phase of our project, we encountered raw transaction data from the market basket analysis data. Data processing was the first important step in our journey. This involved cleaning the data, handling missing values, and converting the data into a format suitable for association analysis. We applied one-time encoding to convert the categorical data into binary format, ensuring that the algorithms we used could interpret and extract meaningful patterns from the data set.

SELECTION OF MODELS:

With our data now in a suitable format, the next important step was model selection. Our project delved into advanced association analysis techniques to extract meaningful insights from the dataset. We used several techniques including FP-Growth, Eclat, Sequential Pattern Mining and Temporal Association Rule Mining. Each technique provides a unique perspective on the data set, and the choice of technique depended on the specific analysis objectives and the nature of the data.

MODEL EXECUTION:

Once the models were selected, we ran them on the preprocessed dataset. This involved applying selected algorithms such as FP-Growth to market basket analysis to find recurring items, association rules and event sequences. We also adjusted the parameters of each technology to optimize their performance and relevance to our project goals.

Association Analysis Techniques:

The core of our project is the application of advanced association analysis techniques. FP-Growth allowed us to efficiently find common sets of items, while Eclat provided an alternative perspective for mining sets of items. Sequential pattern mining was used to uncover sequences of events, providing valuable insight into the temporal nature of our data. In addition, temporal association rule mining allowed us to explore timerelated patterns in the dataset, helping us identify relationships that evolve over time.

PROGRAM:

```
Import pandas as pd
Import matplotlib.pyplot as plt
Import seaborn as sns
From mlxtend.frequent_patterns import fpgrowth,
apriori, association_rules
From prefixspan import PrefixSpan
From temporal_assoc_rules import
TemporalAssocRules
From pyECLAT import ECLAT

# Load your dataset from an XLSX file (replace
'your_dataset.xlsx' with your file)
Data = pd.read_excel('/content/csv.xlsx')
```

```
# Perform FP-Growth
Fpgrowth_result = fpgrowth(data, min_support=0.3,
use_colnames=True)

# Perform Apriori
Apriori_result = apriori(data, min_support=0.3,
use_colnames=True)

# Perform ECLAT
Eclat_result = ECLAT(data, verbose=True)

# Perform PrefixSpan
Prefixspan_result = PrefixSpan(data)

# Perform Temporal Association Rules
Temporal_assoc_rules = TemporalAssocRules(data)

# Visualize the results
Plt. Figure(figsize=(12, 4))
```

```
Plt.subplot(151)
Sns.heatmap(association_rules(fpgrowth_result,
metric="lift", min_threshold=1.0).pivot(
    Index='antecedents', columns='consequents',
values='lift'), annoy=True, fmt='.2f',
cmap='coolwarm')
Plt.title('FP-Growth')
Plt.subplot(152)
Sns.heatmap(association_rules(apriori_result,
metric="lift", min_threshold=1.0).pivot(
    Index='antecedents', columns='consequents',
values='lift'), annot=True, fmt='.2f',
cmap='coolwarm')
Plt.title('Apriori')
Plt.subplot(153)
Sns.heatmap(eclat_result.astype(int),
cmap='coolwarm', cbar=False)
Plt.title('ECLAT')
Plt.subplot(154)
Prefixspan_result.topk(5, closed=True)
```

```
Plt.title('PrefixSpan')
Plt.subplot(155)
Temporal_assoc_rules.find_rules(support_threshold=0.2, time_window=2)
Plt.title('Temporal Assoc Rules')
Plt.tight_layout()
Plt.show()
```

CONCLUSION:

In conclusion, our project demonstrates the power of advanced association analysis techniques to extract meaningful patterns and insights from complex datasets. We carefully processed the data, selected the appropriate models, applied them effectively and studied the material comprehensively. Through these efforts, we have gained valuable insights into shopping cart trends and time patterns in the market that can aid in decision making and strategy development. This project highlights the importance of data-driven analysis and demonstrates the importance of model selection, data pre-processing and dataset exploration to achieve effective results in the field of data science.