WQD 7005 | Tutorial 9

1.

Association rules are a type of rule-based machine learning that identifies relationships between variables in large databases. They are typically expressed as "if-then" statements, such as "if a customer buys diapers, then they are likely to also buy chips." Association rule mining is a subfield of data mining that focuses on finding these types of relationships.

Association rule mining has a wide variety of applications, including

a. Market Basket Analysis: This is the original application of association rule mining, and it is still one of the most common. Market basket analysis is used to identify relationships between products that are frequently purchased together.

b. Fraud Detection: An association rule miner might identify a pattern of transactions where a customer makes a large purchase with a credit card that has never been used before.

c. Customer Segmentation: Association rule mining can be used to segment customers into groups based on their purchase behavior. This information can be used to target marketing campaigns and improve customer service.

There are a number of challenges in the field of association rule mining, including

a. Data Sparsity: Many databases are too sparse to be effectively mined for association rules. This is because the rules are typically only found when two or more items are frequently purchased together.

b. Scalability: Association rule mining algorithms can be computationally expensive, especially for large databases. This can make it difficult to mine large datasets in a timely manner.

c. Noise: Association rule mining algorithms can be sensitive to noise in the data. This means that the rules that are found may not be accurate if the data contains errors or outliers.

2.

Frequent itemsets are sets of items that frequently co-occur together in a transactional or relational dataset. In the context of association rule mining, a frequent itemset refers to a set of items that appear together in a certain proportion of transactions or instances. The goal of association rule mining is to discover interesting relationships or associations between items in a dataset.

Association rule mining relies on the concept of frequent itemsets to identify patterns or associations that occur frequently in the data. It follows a two-step process:

Nasir Uddin Ahmed
Student Id: S2015449

a. Frequent Itemset Mining: This step identifies the itemsets that occur frequently in the data. An itemset is a set of items, and its frequency is the number of times it appears in the data. The frequent itemsets are used to generate association rules.

b. Association Rule Generation: This step generates association rules from the frequent itemsets. An association rule is an "if-then" statement that describes a relationship between two or more items. The rule is typically expressed as "if item A occurs, then item B is likely to occur."

Measures such as support and confidence are used to evaluate the significance of frequent itemsets and association rules

a. Support measures how frequently an itemset appears in the data. It is calculated as the fraction of transactions that contain the itemset. A high support value indicates that the itemset is frequently occurring in the data.

b. Confidence measures how likely it is that item B occurs given that item A occurs. It is calculated as the ratio of the number of transactions that contain both items A and B to the number of transactions that contain item A. A high confidence value indicates that item B is likely to occur if item A occurs.

3.

Integration of association rule mining with other data mining tasks can provide valuable insights and enhance the overall analysis. Here are some ways to integrate association rule mining with other tasks

a. Classification: Association rule mining can be used to discover frequent itemsets in the dataset, and then the extracted rules can be used as features in a classification task. By incorporating the discovered association rules as additional input variables, classifiers can potentially benefit from the discovered patterns and improve their predictive performance.

b. Clustering: Association rules can be used to analyze the relationships between items in different clusters. After clustering the data, association rule mining can be applied within each cluster separately to identify interesting relationships specific to that cluster. This can provide insights into the behavior and characteristics of different clusters.

c. Outlier detection: Association rule mining can help identify unusual or rare itemsets that deviate from the norm. By considering infrequent itemsets or rules as potential outliers, association rule mining can contribute to outlier detection tasks. Outliers can represent anomalies, unexpected patterns, or interesting deviations from the majority of the data.

SAS Enterprise Miner provides functionality for association rule mining through its Association node. Here's an overview of the steps involved in performing association rule mining using SAS Enterprise Miner

Data preparation: Load the dataset into SAS Enterprise Miner and preprocess it as required. Ensure that the data is in the appropriate format for association rule mining, such as transactional or market basket format, where each row represents a transaction or instance.

a. Create a New Project: Create a new project in SAS Enterprise Miner and define the variables to be used for association rule mining.

b. Add the Association Node: Drag and drop the Association node from the toolbox onto the diagram workspace.

c. Configure the Association node: Configure the parameters of the Association node, such as the input data, item selection method, support threshold, and confidence threshold. We can also specify additional options like maximum number of items and maximum rule length.

d. Running the Analysis: Execute the analysis by running the project or specifically the Association node. SAS Enterprise Miner will perform the association rule mining and generate the results.

e. Interpret the Results: Analyze the generated association rules, including their support, confidence, lift, and other relevant measures. Identify interesting and significant rules based on the domain knowledge and objectives of the analysis.

Best practices and tips for using SAS Enterprise Miner for association rule mining include:

a. Understand the Data: Have a thorough understanding of the dataset, its structure, and the context in which association rules are being mined. This understanding will help in selecting appropriate parameters and interpreting the results effectively.

b. Experiment with Different Settings: Adjust the support threshold, confidence threshold, and other parameters to explore different patterns and rule sets. Experimentation can help uncover different levels of detail or interesting relationships.

Nasir Uddin Ahmed
Student Id: S2015449

c. Consider Pre-processing: Depending on the dataset, pre-processing steps like data transformation, filtering, or feature engineering may be beneficial to improve the quality of the results.

d. Validate and Refine the Rules: Perform a comprehensive evaluation of the generated rules using various measures like lift, conviction, or interestingness. Refine the rules by adjusting the thresholds or incorporating domain knowledge.

4.

Association rule mining can be applied in different domains and industries to gain valuable insights from large datasets. Here are some examples of how association rule mining can be applied in different domains

Domains	Name	Description
Retail	Product Recommendations	Association rule mining can be used to recommend products to customers based on their past purchases. This information can be used to improve the customer experience and increase sales.
Retail	Fraud Detection	Association rule mining can be used to identify fraudulent transactions by looking for patterns that are unusual or suspicious. For example, an association rule miner might identify a pattern of transactions where a customer makes a large purchase with a credit card that has never been used before.
Healthcare	Diagnosis	Association rule mining can be used to identify patterns in medical records that can be used to diagnose diseases. For example, an association rule miner might identify a pattern of symptoms that are frequently associated with a particular disease.
Healthcare	Drug Discovery	Association rule mining can be used to identify patterns in drug interactions that can be used to develop new drugs. For example, an association rule miner might identify a pattern of drugs that are frequently prescribed together that could be used to develop a new drug that combines the effects of the two drugs.
Finance	Risk Assessment	Association rule mining can be used to assess the risk of default on loans or other financial instruments. For example, an association rule miner might identify a pattern of financial data that is associated with a high risk of default.

Finance	Customer Segmentation	Association rule mining can be used to segment customers into groups based on their financial behavior. This information can be used to target marketing campaigns and improve customer service.
Marketing	Product Recommendations	Association rule mining can be used to recommend products to customers based on their past purchases. This information can be used to improve the customer experience and increase sales.
Marketing	Campaign Optimization	Association rule mining can be used to optimize marketing campaigns by identifying the most effective targeting criteria and offers.

Examples of successful applications of association rule mining:

a. Amazon: Amazon uses association rule mining to recommend products to customers. For example, if a customer buys a book on Java programming, Amazon might recommend other books on Java programming, as well as books on related topics such as software development or computer science.

b. Walmart: Walmart uses association rule mining to identify products that are frequently purchased together. This information is used to improve product placement in stores and to target marketing campaigns.

c. Netflix: Netflix uses association rule mining to recommend movies and TV shows to customers. For example, if a customer watches a movie about a superhero, Netflix might recommend other movies about superheroes, as well as movies in the same genre, such as action movies or science fiction movies.

5.

Evaluating and comparing different sets of association rules can help identify the most interesting and relevant rules. Here are some techniques to evaluate and compare association rules:

a. Pruning: Pruning involves removing redundant or uninteresting rules from the rule set. Redundant rules are those that convey the same information as other rules. Pruning techniques can be applied based on measures like confidence, support, or lift to eliminate rules that do not provide additional insights.

b. Filtering: Filtering allows us to focus on a subset of rules that meet specific criteria or conditions. For example, we can filter rules based on minimum support or confidence thresholds, minimum itemset size, or specific items of interest. Filtering helps to reduce the number of rules and narrow down the focus to the most relevant ones.

c. Ranking: Ranking involves ordering the rules based on a particular measure or scoring criterion. Common ranking measures include lift, confidence, support, or interestingness measures such as the Gini index or leverage. By ranking the rules, we can identify the top rules that are most significant or interesting according to the chosen measure.

d. Evaluation Measures: Various evaluation measures can be used to assess the quality or interestingness of association rules. Some commonly used measures include lift, confidence, support, conviction, and interestingness measures like the chi-square statistic or information gain. By considering multiple evaluation measures, we can gain a comprehensive understanding of the rules' significance.

SAS Enterprise Miner is a data mining software that can be used to perform these tasks. SAS Enterprise Miner provides a variety of tools for evaluating and comparing association rules. These tools include:

a. The Association Rules Viewer: The Association Rules Viewer is a graphical tool that can be used to visualize association rules. The viewer allows us to see the support, confidence, and lift of each rule, as well as the items that are included in the rule.

b. The Association Rules Analyzer: The Association Rules Analyzer is a statistical tool that can be used to analyze association rules. The analyzer allows us to calculate various metrics for each rule, such as support, confidence, lift, and conviction.

c. The Association Rules Profiler: The Association Rules Profiler is a tool that can be used to profile association rules. The profiler allows us to see how the rules change as we vary the support and confidence thresholds.