**Tutorial 01 Answers**

1. What are the two major aspects related to data science. Briefly explain the focus of each related aspect.

* The management and processing of data
* The analytical methods and theories for data analysis and optimization

1. What are the three major types of analytic techniques related to data science? Briefly explain for what kinds of purposes each analytic techniques are used for, techniques used in each type of analytic technique and one practical application of the analytic technique.

* Descriptive analytics

**Purposes used**: primary purpose is to summarize and understand existing

data.

**Techniques used: data measurement technique**

**Practical example:** commonly used in bar charts, box plots, and

Scatter plots

* Predictive analytics

**Purposes used**: extracting information from data sets to determine patterns and predict future outcomes and trends.

**Techniques used: \*** Classification techniques

\*Regression techniques

\*Time series models

**Practical example: \*** Credit card, banking and finan­cial services- To detect and reduce fraud, measure credit risk.

* Media and entertainment- Deepen insight into audiences by identifying influencing attributes, trends, drivers and desires.
* Prescriptive analysis

**Purposes used**: provide recommendations in support of decision-making processes, where the objective is to determine a set of decisions and/or actions that gives rise to the best possible results subject to various constraints.

1. What is meant by an association rule? When is an association rule considered strong?

* Association rule is which aims to observe frequently occurring patterns , corrections or associated data set found in various kind of databases.
* Association rule support confidence greater than or equal to a user specified minimum support threshold.

1. From a business perspective how could identifying association rules be helpful to improve sales?

* Identifying association rules can help in finding frequently occurring patterns and correction within a business transaction. This would help forecast your business and take appropriate decisions to improve the sales based on those forecasts.

1. What are the two steps in association rule mining?

* Finding all frequent item sets.
* From all the frequent item sets, generate association rules satisfying the minimum support conditions.

1. Consider the following set of transactions. Assuming that the minimum support is 60%, identify the frequent itemsets within the set of transactions.

|  |  |
| --- | --- |
| TID | Items |
| T1 | Butter, bread, milk, sugar |
| T2 | Butter, flour, milk, sugar |
| T3 | Butter, eggs, milk, salt |
| T4 | eggs |
| T5 |  |

**Item set 1**

|  |  |
| --- | --- |
| Items | support |
| Butter | 4 |
| Bread | 1 |
| Milk | 4 |
| sugar | 3 |
| Flour | 2 |
| Eggs | 2 |
| salt | 2 |

**Item set 2**

|  |  |
| --- | --- |
| Items | support |
| Butter,milk,sugar | 3 |

**Item set3**

Let’s assume;

**S= 𝜎**( Butter. Milk, sugar)

**=3/5 = 60%**

**|T|**

1. 7. Consider the following set of transactions. Assuming that the minimum support is 33.3%, identify the frequent itemsets within the set of transactions.

|  |  |
| --- | --- |
| TID | Items |
| T1 | Hotdogs, Buns, Ketchup |
| T2 | Hotdogs, Buns |
| T3 | Hotdogs, Coke, Chips |
| T4 | Chips, Coke |
| T5 | Chips, Ketchup |
| T6 | Hotdogs, Coke, Chips |

1. **Item set**

|  |  |
| --- | --- |
| Items | support |
| Hotdog | *4* |
| Buns | 2 |
| Ketchup | 2 |
| Coke | 3 |
| Chips | 4 |

1. **Item set**

|  |  |
| --- | --- |
| **Item** | **support** |
| Hotdog, Chips, coke | **3** |

**Item set- 3**

**Let us assume:**

**S = σ**( Hotdogs, chips, coke) = 2/6 =33.3%

|T|

Then let us assume;

**C = σ**( Hotdogs, chips, coke) = 2/2 =100%

**σ** (Hotdog, chips)

Therefore, there is a strong rule