



School of Computer Science and Engineering

LAB - 5 & 6 Exercises

Course Code	:	CSE6017 – Mining Massive Dataset	Date	:	17/03/2021 and 24/03/2021
Lab Experiment	:	Apriori Algorithm in Apache Spark	Slots	:	L45+L46
Instructors	:	Dr. Ramesh Ragala			

Objective:

To understand and implement the Apriori Algorithm for Association Rule mining in Apache Spark framework.

<https://towardsdatascience.com/big-data-market-basket-analysis-with-apriori-algorithm-on-spark-9ab094b5ac2c>

The content in the above link describes the implementation of Apriori Algorithm in Apache Spark from scratch. Fruit.txt dataset is used in that program.

1. In the practical exercise, you need to implement the Apriori algorithm based on the above link with **strong association rules** with **minimum confidence** as 75%. And also find the time takes to complete the task.
2. In the practical exercise, you need to implement problem given in (1) using **Data Frames** (One of the data structure in Apache Spark)