



School of Computer Science and Engineering

LAB - 3 & 4 Exercises

| | | | | | |
|----------------|---|---|-------|---|----------------------------------|
| Course Code | : | CSE6017 – Mining Massive Dataset | Date | : | 17/02/2021 and 24/02/2021 |
| Lab Experiment | : | Decision Tree in Apache Spark | Slots | : | L45+L46 |
| Instructors | : | Dr. Ramesh Ragala | | | |

Objective:

To understand and implement the tree learning methods (decision tree) in Apache Spark framework.

1. Implement the Decision tree (for classification based on gini index) in Apache Spark using pipeline and MLlib packages on IRIS dataset. And also implement the same from the scartch without using any built-in functions. Finally compare the Acurracy Results.
2. Implement the Decision tree (for classification based on Entropy) in Apache Spark using pipeline and MLlib packages on IRIS dataset. And also implement the same from the scartch without using any built-in functions. Finally compare the Acurracy Results.
3. Implement the Decision tree (for Regression based on Variance) in Apache Spark using pipeline and MLlib packages on IRIS dataset. And also implement the same from the scartch without using any built-in functions. Finally compare the Acurracy Results.