



VIT[®]

Vellore Institute of Technology

(Deemed to be University under section 3 of UGC Act, 1956)

School of Computing Science and Engineering

LAB - 6 Exercises

Course Code	:	CSE6017 - Mining Massive Dataset	Date	:	14/02/2020
Lab Experiment	:	Decision Tree algorithm implementation in Apache Spark framework	Slots	:	L55+L56
Instructors	:	Prof. Ramesh Ragala			

Objective:

1. To understand the Decision Tree Implementation in Apache Spark using PySpark

Exercises:

1. Use the following to know the implementation details of Decision tree Algorithm using ML in PySpark. Use the OneHotEncoder, Pipeline concept, etc.

<https://towardsdatascience.com/machine-learning-with-pyspark-and-mllib-solving-a-binary-classification-problem-96396065d2aa>