



**VIT<sup>®</sup>**

**Vellore Institute of Technology**

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

## **School of Computing Science and Engineering**

### **LAB - 7 Exercises**

<b>Course Code</b>	<b>:</b>	<b>CSE6017 - Mining Massive Dataset</b>	<b>Date</b>	<b>:</b>	<b>05/03/2020</b>
<b>Lab Experiment</b>	<b>:</b>	<b>K-Means algorithm implementation in Apache Spark framework</b>	<b>Slots</b>	<b>:</b>	<b>L55+L56</b>
<b>Instructors</b>	<b>:</b>	<b>Prof. Ramesh Ragala</b>			

Objective:

1. To understand the K-Means algorithm implementation in Apache Spark using PySpark

Exercises:

1. Perform the pre-processing on iris dataset and check the performance of K-means clustering algorithm [i.e normalize the data in the range of (-1,1)]
2. Implement K-Means++ on the uber.csv dataset.