WEEKLY PROGRESS REPORT

Directed Research Projects

Zohaib Ahmad & Waleed Irfan
Student ID : 170902110 & 170406830
ahma2110@mylaurier.ca &
irfa6830@mylaurier.ca

Immediate Supervisor : Dr. Safaa Bedawi
Assistant : Professor Sukhjit Singh Sehra

January 2021

WEEKLY REPORT #1
DIRECTED PROJECTS

Student Names : Zohaib Ahmad & Waleed Irfan
Department: Computer Science (Honors)
Date: January 22, 2021
Immediate Supervisor: Dr. Safaa Bedawi
Assistant : Professor Sukhjit Singh Sehra

TASK	STATUS	DURATION
Working on Hadoop Mini Project	Finished uploading csv to Hadoop, configuring features now	1 week

WORK DONE

First Meeting (Tuesday Jan 19)
Decide on first call for teams (Friday Jan 22)

Exported Data from Kaggle to csv file

1) Installed Hadoop
2) Imported csv file

Student ID : 170902110 & 170406830

Hadoop

- Apache Hadoop is a open source software that of which uses network of many computers to solve problems that involve massive amounts of data and computation
  - Developed by Google, Apache Software Foundation, MapR

- It works by distributing processes for huge data sets across cluster of commodity servers and works on multiple machines simultaneously and requires the program "Hadoop" to process any data, HDFS stores data while MapReduce process data and Yarn divides Task

- For directed research project, Hadoop will be used to store the kaggle problem "Alexa ranking" in order to help process information for an algorithm later developed by both students

- This week both students have successfully installed the required programs to have Hadoop running and achieved storing the csv file locally.

- Immediate goals goals going into next week are to organize Hadoop to maximize SEO and to decide what features will be most useful in algorithm

As this was our first time installing Hadoop on our machines, we followed a well structured video on YouTube which guided us on how to install Hadoop on our machines. This included steps such as configuring our Java version on our machine, installing and configuring the Hadoop files, making sure it is properly integrated with our system, and eventually deploying it.

References :

Big Data Analytics with R and Hadoop, By : Prajapati Vignesh, eText ISBN : 9781782163299,1782163298

https://www.youtube.com/watch?v=26w6IL8_7Q8&t=1302s

https://httpd.apache.org/docs/2.4/platform/windows.html