TED UNIVERSITY

Faculty of Engineering

Department of Computer Engineering

CMPE 491 (Senior Project 1) – Project Proposal

by

Yağız ÇİMEN – Efe TONTU – Mesut Nadir SEYFELİOĞLU

Supervisor: Assoc. Prof. Dr. Tansel DÖKEROĞLU

Name of the Project:

- Name: EBIG

The URL of the project web page:

- url: https://github.com/mnadirs/e-big

Names of the team members:

- Yağız ÇİMEN
- Mesut Nadir SEYFELİOĞLU
- Efe TONTU

Names of the Supervisor and Jury Members:

- Tansel DÖKEROĞLU (Supervisor)
- Venera ADANOVA (Jury Member)
- Emin KUĞU (Jury Member)

Description of the Project:

Aim of this Senior Project is to study big data datasets in order to analyse the society communities in social media platform(s). Nowadays lots of company prefers advanced, faster communication platforms like social media to advertise their products and reach out more customers. As more customers intrested in products, their profit will increase in right proportion. Our team will use this information from big data researchs for the purpose of provide knowledge to those firm who wants to specify their advertisement fields. Simultaneously we can achieve the transformation of data to value which is one of the most important things today.

In this project we are going to use the youtube, instagram and twitter datasets. Our aim is to analyse the statistics of trending youtube videos and comments. First we going to use Cloudera on Docker. We will manage our big data in Hadoop HDFS system. For the analysis part we are going to use Apache Spark and its libraries. Also MLlibs in spark will help us to analyse the data. Our aim is to implement deep learning algorithms to our project. In addition to Java, Python Technologies will be used. Lastly we planned to transform our analyse to a mobile app as a product. React and other mobile techs. are going to be used.

Datasets:

-https://www.kaggle.com/datasnaek/youtube?select=GBvideos.csv

-https://www.kaggle.com/datasnaek/youtube-new?select=CAvideos.csv

-https://www.kaggle.com/mathurinache/mostinstagramfollowers (experimental)

-https://www.kaggle.com/crowdflower/twitter-user-gender-classification (experimental)