CSEE 5590/490: Big Data Programming

Instructor - Zeenat Tariq

Team Number :- 6

Project Proposal (Increment 1)

1. Project Title and Team Members

International Student Data Analysis using Twitter API
Team Members :-
1) Tanvi Jain
2) Saikumar Reddy Papagari
3) Amarnadha Reddy Ankireddypalli
4) Thotakura Naga Mounika

2. Goals and Objectives:

- Motivation

The members of this team belong to the international student community and we came to the US to get a higher degree. There are many instances/scenarios where we are expected to follow those norms which the general US citizens are not expected to follow eventually because we belong to the immigrant group. This gave us the foundation for this project and we decided to showcase the essential data like percent of students coming to the United States for education, average amount of money spent on education, probability of getting a work VISA, the majors that mostly students like to get enrolled into and study from the US.

Also we wanted to address the problems/privileges that an international student comes across due to the immigrant rules

and what impact it has on a student who belongs from a foriegn land.

This study has never been taken up before especially in the form of a project and we only have blurred vision about the percentages and figures and these above listed queries, so we wanted to take up the challenge and showcase these crucial discussions by the help of Twitter API and big data programming and visualization tools.

- Significance

  There are currently over 1 million international students from more than 220 countries, coming to the US annually. This project will have a significant impact on the prospective future students and will help them in making the right choice by giving a clearer vision of the essential discussions. The project analyses the major challenges that the international students are currently facing like cultural shocks, financial crises and lack of job opportunities due the work visa requirements and immigration rules.

- Objectives

  The objective of this project is to get an analysis of the student data to provide a clearer view of the challenges that international students face.

  The analysis of the data is done by extracting the datasets from the Twitter streaming API and the datasets are executed using Hadoop/Spark and at the end data is visualized using Tableau

- Features

  The main features of this project is the twitter data extraction through twitter streaming API, and using

NLP/NLTK for sentiment analysis on the extracted data. We will also use visualization tools like Tableau or python integrated libraries for revealing the essential details.

3. Increment 1 storytelling:

1. Who are the people or communities in need of help?
   The international students who are studying in the United States Of America.(or overseas)

2. What problem happened to them?
   The problem mainly affected them in terms of finance, job hunting, allowance to work with any company at any time or lack of getting accepted by any company due to limited work authorization and stress because of lack of stability in life.

3. When did the problem take place?
   There is no specific time/place of problem, it is the general situational challenges that international students come across.

4. Where means two things:
   These challenges can be faced by international students in a foriegn land, where they are a part of the immigrant community and do not have access to the perks that the citizens of that country get, which makes life a bit harder for even the very basic requirements of living.

5. Why means the possible causes and/origin of the problem.
   Due to a set of immigration rules and less availability of jobs/ work authorization in desired fields.

6. How: If you would like, you can add a dimension of how. How did it happen? Sometimes, the answer to how can be covered by what, when and where.
   There isn't a certain time that the chances of happenings like these because these kind of incidents are seen and

experienced by many international students over the period of time.

4. Contribution of Work

Since this is the initial phase of the project, the contribution is not clear at this point, but the work will be divided among the team members as follows :-

1) Tanvi Jain & Thotakura Naga Mounika
   Extracting the data related to international students and reaction tweets to the immigration rule changes through Twitter Streaming API and organizing the data of later processes.
2) Saikumar Reddy Papagari & Amarnadha Reddy Ankireddypalli
   Exploring the data and BDP tools on how to process the extracted data and filter it using Spark/hadoop or other BDP frameworks to infer highlights and crucial information about the subject.

5. References

- https://towardsdatascience.com/how-to-build-a-real-time-twitter-analysis-using-big-data-tools-dd2240946e64
- https://medium.com/@jaimezornoza/downloading-data-from-twitter-using-the-streaming-api-3ac6766ba96c