

Analytics for FAST Alumni Network

Project Team

Basharat Hussain p17-6102
Tahir Aziz p17-6054

Session 2017-2021

Supervised by

Mr. Muhammad Amin



Department of Computer Science

**National University of Computer and Emerging Sciences
Peshawar, Pakistan**

July, 2021

Student's Declaration

We declare that this project titled “*Analytics for FAST Alumni Network*”, submitted as requirement for the award of degree of Bachelors in Computer Science, does not contain any material previously submitted for a degree in any university; and that to the best of our knowledge, it does not contain any materials previously published or written by another person except where due reference is made in the text.

We understand that the management of Department of Computer Science, National University of Computer and Emerging Sciences, has a zero tolerance policy towards plagiarism. Therefore, We, as authors of the above-mentioned thesis, solemnly declare that no portion of our thesis has been plagiarized and any material used in the thesis from other sources is properly referenced.

We further understand that if we are found guilty of any form of plagiarism in the thesis work even after graduation, the University reserves the right to revoke our BS degree.

Basharat Hussain

Signature: _____

Tahir Aziz

Signature: _____

Verified by Plagiarism Cell Officer

Dated:

Certificate of Approval



The Department of Computer Science, National University of Computer and Emerging Sciences, accepts this thesis titled *Analytics for FAST Alumni Network*, submitted by Basharat Hussain (p17-6102), and Tahir Aziz (p17-6054), in its current form, and it is satisfying the dissertation requirements for the award of Bachelors Degree in Computer Science.

Supervisor

Mr. Muhammad Amin

Signature: _____

Ms Mashal Khan

FYP Coordinator

National University of Computer and Emerging Sciences, Peshawar

Dr. Hafeez

HoD of Department of Computer Science

National University of Computer and Emerging Sciences

Acknowledgements

In the name of Allah, the Most Gracious and the Most Merciful. We have completed our Final Year Project Part 1. With the grace of Allah, We are at a level where we can present prototype of our project. Inshallah in next semester we are planning to vastly improve our project and add more features in our project. In our project we have faced many difficulties and hardships. In these difficulties our supervisor Mr Muhammad Amin has helped us alot by guiding us throughout the project. Checking our project and showing us improvements that need to be done and mistakes that need to be sorted out. With that giving us an opportunity to learn and experiment throughout practical work.

Basharat Hussain

Tahir Aziz

Abstract

In times of increasing universities graduates and competitive job markets. It's difficult for an institute to know what's happening in job markets and how their Alumni are performing. Not knowing the statistics can create false assumptions and can be misleading. Finding facts and figures can be tricky and time consuming. Techniques like Data Scraping and Data Mining are used in the process. Going further into data scraping like linkedIn is challenging ,tricky and most of the time failure. By doing analytics for FAST alumni we have achieved to create a broad vision for the institute to better understand the Job market, Alumni contribution and the skill set they need to offer to students in upcoming years.

Contents

1	Introduction	1
1.1	Background	1
1.1.1	Big Data	2
1.1.2	Data Collection	2
1.1.2.1	Data Scraping	2
1.1.3	Data Analysis	3
1.1.4	Data Analytics	4
1.1.4.1	Data Visualization	4
2	Motivation And Scope	5
2.0.1	Motivation	5
2.0.2	Scope	5
3	Work Flow And Project	7
3.0.1	Introduction	7
3.0.2	Difficulties and Problems	7
3.0.3	Web Scraper	7
3.0.4	Data Analysis	8
3.0.5	Data Analytics	8
3.0.6	Flow Chart Explained	8
3.0.7	About Data	10
3.1	Literature Review and Findings	10
3.1.1	Use Case Diagram	12
3.1.1.1	Use Cases	12
4	Conclusions and Future Work	15

4.1	Conclusion	15
4.1.1	Breakdown for Part 1	15
4.1.2	Timeline Milestone	15
4.2	Future Plan for Part 2	16
4.2.1	Future Milestone	16
References		17

List of Figures

3.1	Example Data visualization of Our Project	8
3.2	Flow Chart of Our Project	9
3.3	Progress of Our Project	10
3.4	Use Case Diagram of Our Project	13
4.1	Timeline Milestone of Our Project	15
4.2	Future Timeline of our Project	16

List of Tables

3.1 Literature Review Table. 11

Chapter 1

Introduction

We all know Facebook, Google and Amazon are big tech companies increasing their revenue each year. Amazon as a marketplace for customers to buy products is launching its own product and topping the chart as top seller in the list. These all big companies have one thing in common. They have big databases. They are using big data to their advantage and doing analytics on the data. They can easily find the trends and launch products or services according to it. They double their profits and prepare for any surge they can see in products, services or servers.

In this project, we are using analytics and big data to our advantage. We are collecting data of alumni and doing analytics on it. We need to find how our alumni are performing in the marketplace and how we can improve students' skill set to better compete in the marketplace.

[?]

1.1 Background

Analytics is mostly used for business insight. Analytics can be found as early as the 19th century. In the 1960s when computers were starting to be used in industries. Analytics was starting to appear in many forms such as in industrial management, cost analysis and more.

1.1.1 Big Data

Big data is referred to as a large data set. Big data can be structured and unstructured. Big data is usually created automatically by the customers or providing service. These data can be analysed and patterns can be found. Big data can help in future prediction in any aspect by analyzing past data.

Big data can help businesses in cost reduction, time reduction, new product development, and maximize profits. In our case, students and teachers can get insight into the market-place.

1.1.2 Data Collection

Data collection is the sole need for analysis or analytics [1]. Data collection can be done through google forms or can be taken from websites. Sometimes a data set is available and sometimes you have to scrape the website. There are a lot of ways to scrape a website to extract useful information for reuse. In simple form, we can copy paste snippets from a web page, but this is very difficult and time consuming when there is a lot of data that need to be scraped and web pages. Instead programmers have built different tools and scripts in different programming languages. These scripts navigate themselves to different pages through hyperlinks and extract the data told to and stop when the all pages are searched. to get themselves around websites automatically and extract useful data. Automating web scraping also allows programmers to define whether the script should be run at regular intervals and capture changes in the data.

1.1.2.1 Data Scraping

Data scraping is refer to as automated program that gets data from another program. By data scraping we mostly mean web scraping. Web Scraping saves us lot of time and gives us useful data.

Python community has given us pretty powerful web scraping tools. These are library that can be called in python and used.

[a.]Selenium

Selenium is a browser automation tool that builds to give inputs like humans. It works on various browsers like Firefox, Chrome, Internet Explorer, Edge, Safari. Selenium Webdriver are made for each browser separately which can be called in python script. Selenium is used for automating testing, screenshots in png, get website cookies, change window size to name a few. Selenium is used for scraping , logins, submitting forms, handling alert prompts, adding/deleting cookies, and much more.

[b.]BeautifulSoup

Beautiful Soup is a Python library which is used to get parse tree from HTML, XML, and other markup languages. BeautifulSoup library can be use to extract data from web page which doesn't allow to download data and there is also trash data in between. you can take data that is useful to you by using BeautifulSoup.

1.1.3 Data Analysis

Data analysis is the process of cleaning, interpreting, analyzing and visualizing data. Data analysis is used to discover valuable insights that allow smart decisions for business or institute. Data analysis not only includes data analysis itself, but also data collection, organization, storage and tools and techniques used to filter deep into data. After techniques applied this data shows results on the dashboard - for example, data visualization tools. Data visualization is discussed below in detail.

Data analysis turns raw data into useful statistics, information, prediction or explanations. Relying on intuition or experience to improve services or products -Data analysis provides solid evidence to support decisions that can help institutes or businesses improve these aspects. As well as their overall brand image and customer service[5]. Some of common analysis.

a. Text Analysis

b. Descriptive Analysis

- c. Inferential Analysis
- d. Diagnostic Analysis
- e. Predictive Analysis
- f. Prescriptive Analysis

1.1.4 Data Analytics

Mostly data analysis and data analytics are similar[1]. Data analysis is working on raw data and applying different processes and techniques on data. While data analytics is taking that data and presenting it in such a way that helps in decision making.

1.1.4.1 Data Visualization

Data visualization is the presentation of statically data in a pictorial or graphical format[3].It enables viewers to see analytics presented visually, compare data values with each other and can grasp difficult concepts or identify new patterns with ease. With interactive visualization, you can take the concept a step further by using scripts like pi-chart to drill down into charts and graphs for more detail, interactively changing what data you see and how it's processed.

Chapter 2

Motivation And Scope

2.0.1 Motivation

We listen to our seniors talk about their seniors who have graduated and are doing good jobs. We heard these stories but we couldn't hear for all our seniors graduates. What are they doing? Are they leading the market or are they struggling? To solve this we need data. We need accurate data to compare with other institutes and see the situation in the marketplace.

2.0.2 Scope

We started a project to get answers to the questions in mind of each person in the institute. What alumni have achieved?, where are we standing in the global market in competition and what is the future in the global market. These answers can't be given without data. Data is not available in single form, it needs to be scraped and brought together to be processed and machine learning applied.

Our major constraint is the getting data set[4]. We need to have a lot of data. Without big data we can't have the result we expect. Our main source of data is linkedin as people usually post about their job , skill set, years in company, geolocation and many details like that. Mostly these details are accurate and can be used. Escaping from the ban of linkedin while scraping it is challenging but we have found some tricks that are discussed

in chapter 3 in detail. We are planning to implement machine learning after we extract data to predict the future job market, skillset and compare our alumni with other institutes.

.

Chapter 3

Work Flow And Project

3.0.1 Introduction

Scraping data from linkedin and preprocessing it. Applying techniques to get useful information, predicting top skillset and finding similar patterns unknown to us at the moment.

3.0.2 Difficulties and Problems

When you are doing your project, there can be many difficulties and problems you have to tackle. some of them we have discussed below.

- Web Scrapper
- Data Analysis
- Data Analytics

3.0.3 Web Scrapper

We have used many ready-made scraper to scrape data from linkedin[2]. But problem with these was the data was not related to our alumni and most of it was random data. The scrapper required a premium and most of the linkedin accounts we provided were blocked. So after research and finding. We settled on python libraries beautifulsoup and

selenium. Selenium helped us automate like humans and beautifulsoup let us parse tree from HTML and we could save data with us.

3.0.4 Data Analysis

After getting data from beautifulsoup python library into python script. Still the data was raw and needed preprocessing. We handled the values that were null. like if the alumni is not doing a job. His job section is empty at the moment. We also checked the validity of data and any fake or miss entry data was checked and corrected.

3.0.5 Data Analytics

After getting all the data and preprocessing it. We needed to show our data in beautiful visualisation so the view can understand and relate data with each other. We categorized and classification. here is some example from our visualization.

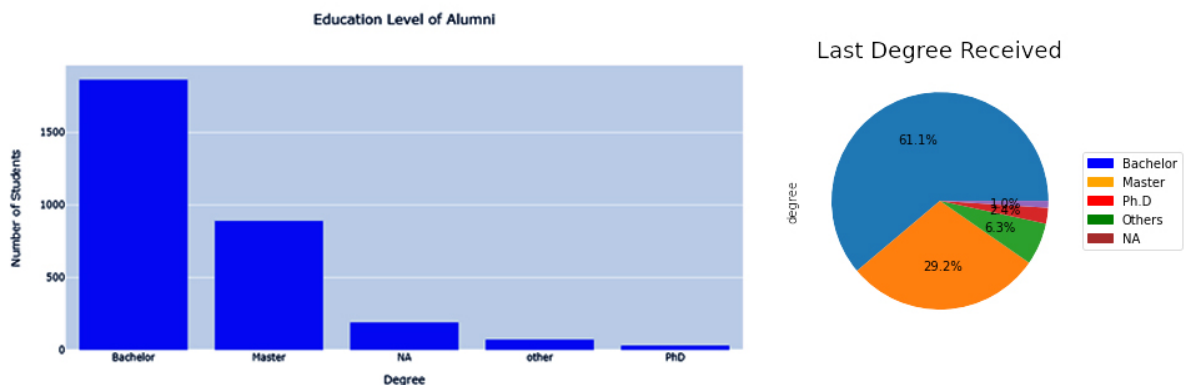


Figure 3.1: Example Data visualization of Our Project

3.0.6 Flow Chart Explained

We are going to get our data from linkedin by web scraping[2]. In web scraping we have to target FAST alumni and scrape the accounts url-id in the first phase. We scrape from recommendations profiles in the search page of linkedin. In the second phase, we have to go to each linkedin id and scrape data from each profile.

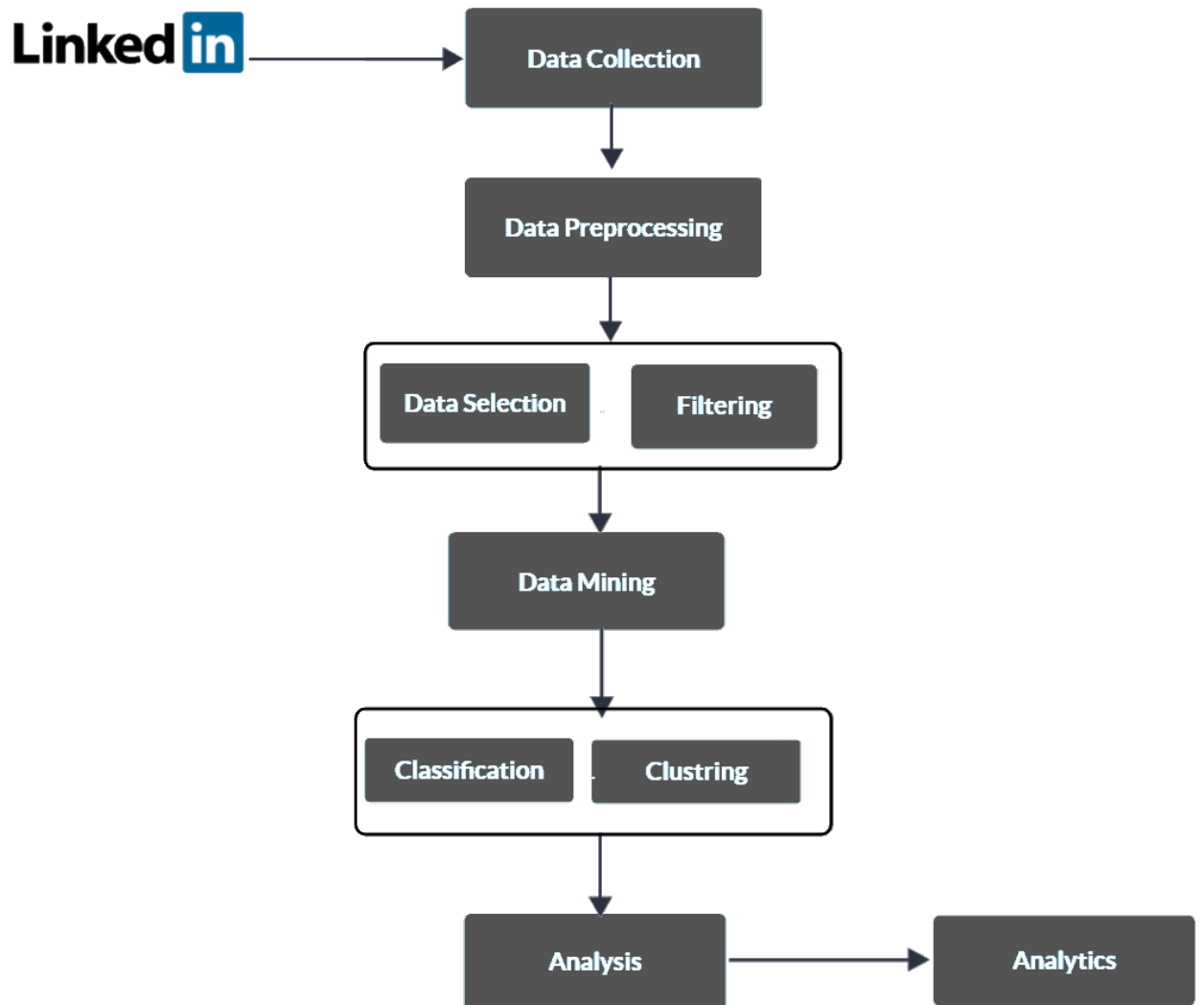


Figure 3.2: Flow Chart of Our Project

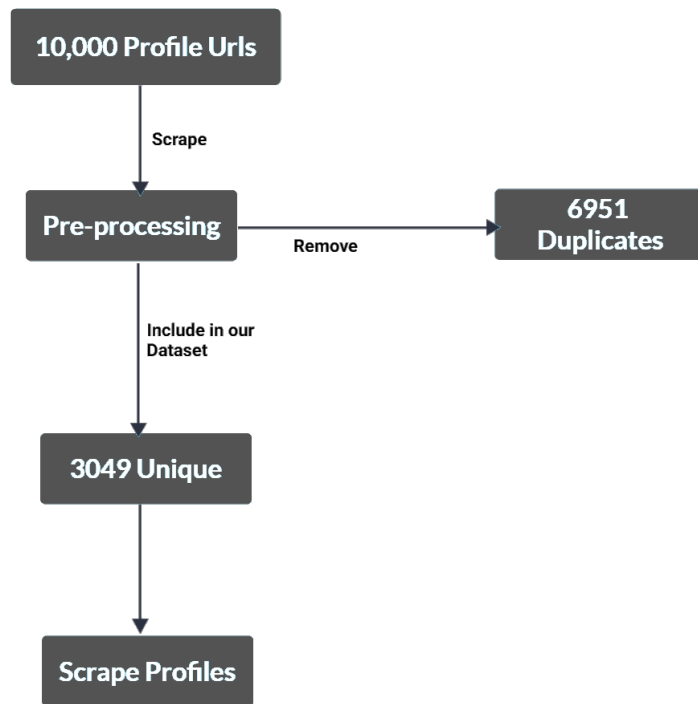


Figure 3.3: Progress of Our Project

3.0.7 About Data

We are successful in scraping 3049 profiles. we have scraped more than 10 thousand profiles. Due to duplicate profiles id we couldn't use all of them.

3.1 Literature Review and Findings

We have studied many research paper and literature to help us in product and here are the findings from them. given in table below.

Sr. No.s	Author Name	Title	Findings
[1]	Chakraborty, Anwesha, et al	Application of graph theory in social media	Various aspects of graphs and its basic properties. Graphs are used in social networks. Complexity is increasing with the advancement in social media.
[2]	Rubens, Neil, et al	Alumni network analysis	Role of alumni play in the reputation of educational institutions
[3]	Serrat, Olivier..	Social network analysis	SNA and its properties
[4]	Markovska, Veneta, and Stanimir Kabaivanov	Improving alumni network efficiency with machine learning	SVM and K-means clustering algorithms. ML algorithms can help a lot in maintaining alumni networks.

Table 3.1: Literature Review Table.

3.1.1 Use Case Diagram

we have made use case diagram to better understand our project.

3.1.1.1 Use Cases

- Alumni can create accounts on linkedIn.
- They can enter their jobs title.
- They can enter a job location.
- They can enter their degree.
- They can enter their degree.
- They can enter more information degrading their career.
- That data is accessible to everyone.
- That data can be preprocessed.
- That data can be analyzed.
- That data analytics can be done.

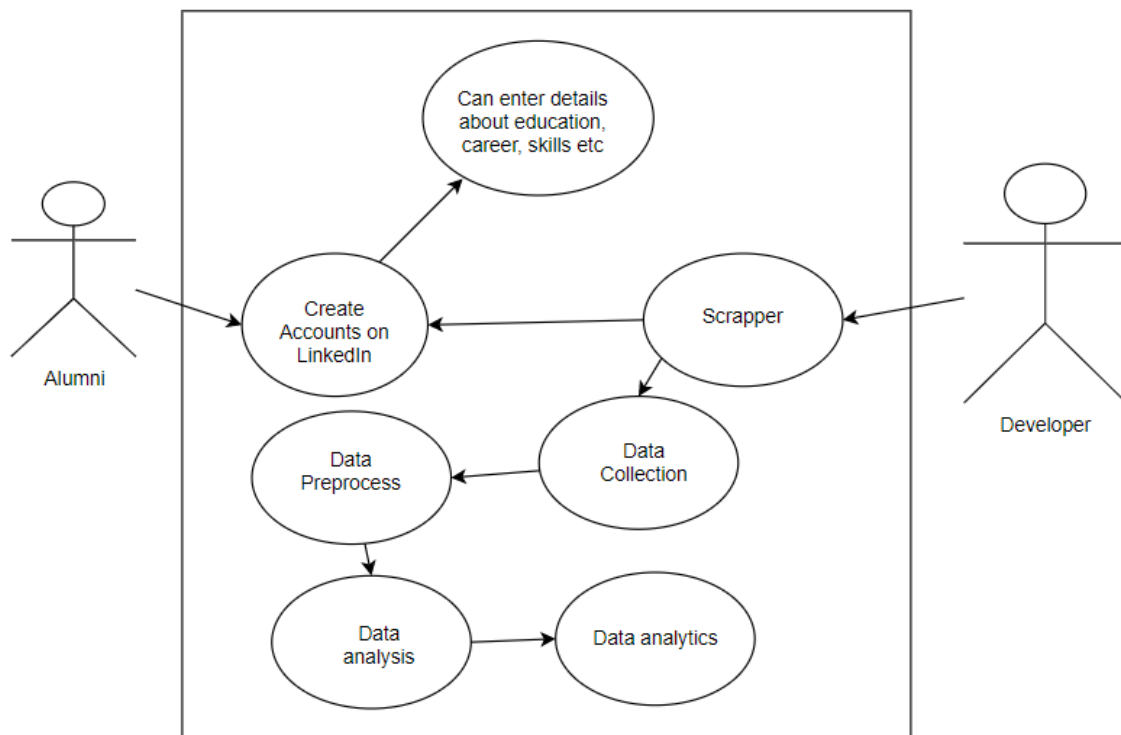


Figure 3.4: Use Case Diagram of Our Project

Chapter 4

Conclusions and Future Work

4.1 Conclusion

We have completed Final Year Project Part 1. We have build prototype Alumni dashboard that displays visual data about our alumni.

4.1.1 Breakdown for Part 1

We researched about scraping and successfully scraped data from linkedin. We have been successful to preprocess and do analysis on the data. We have displayed that data.

4.1.2 Timeline Milestone

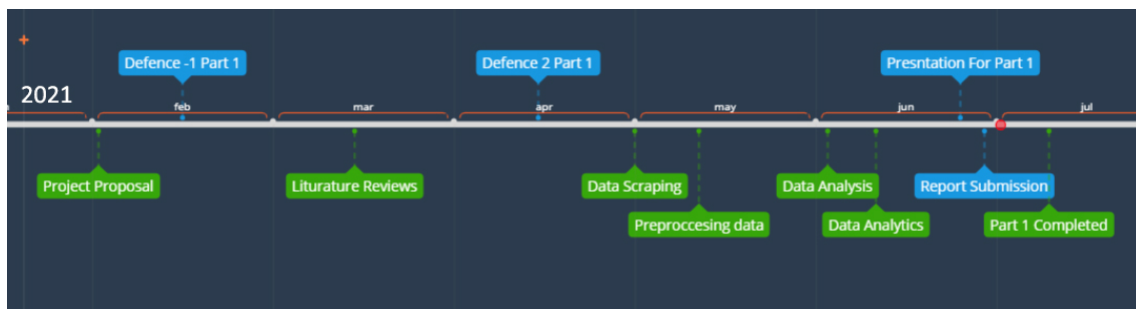


Figure 4.1: Timeline Milestone of Our Project

4.2 Future Plan for Part 2

In final year part 2, we are planning to make improvements in analysis. We will be focusing on educational analysis and improving it so undergraduate students can benefit from it. We will be adding geolocation, job titles, skillset, trending skills and alumni skills comparison. We will be comparing our alumni in Pakistan and the global marketplace. We are also planning to implement machine learning to predict future skill sets and possible jobs in the marketplace. Clustering and classification using Machine Learning. We will also be finding hidden patterns in our data. And lastly we will try to implement new scraping techniques and try to scrape as much data as possible.

4.2.1 Future Milestone



Figure 4.2: Future Timeline of our Project

Bibliography

- [1] et al Chakraborty, Anwesha. Application of graph theory in social media. In *International Journal of Computer Sciences and Engineering*, pages 722–729. IEEE, 2011.
- [2] Kais Dai, Celia González Nespereira, Ana Fernández Vilas, and Rebeca P Díaz Redondo. Scraping and clustering techniques for the characterization of linkedin profiles. *arXiv preprint arXiv:1505.00989*, 2015.
- [3] Michael M Orozco. College of computer studies graduate tracer with data analytics. In *Abstract Proceedings International Scholars Conference*, volume 6, pages 220–220, 2018.
- [4] Olivier. Serrat. Career trajectory analysis of information technology alumni: A linkedin perspective. In *Proceedings of the 17th Annual Conference on Information Technology Education*, pages 2–6, 2016.
- [5] Hapnes Toba, Evelyn A Wijaya, Maresha C Wijanto, and Oscar Karnalim. Developing alumni job prediction models based on term occurrences and word class analysis of search engine page results. *World Transactions on Engineering and Technology Education*, 14(4):563–567, 2016.