Retrieving and Classifying LinkedIn Job Titles for Alumni Career Analysis

Lei Li, Svetlana Peltsverger,
Jack Zheng, Linh Le
Kennesaw State University
680 Arntson Drive
Marietta, GA 30060
001-470-578-3915
{lli, speltsve, gzheng, lle13} @kennesaw.edu

Michael Handlin
Kennesaw State University
680 Arntson Drive
Marietta, GA 30060
001-470-578-5036
mhandlin@students.kennesaw.edu

ABSTRACT

It is important for universities to track alumni careers after graduation for both program quality improvement and accreditation requirements. The most common way of collecting such data is through alumni surveys, but historically these surveys had low response rates. LinkedIn, a social network site that targets working professionals, provides a great platform for academic programs to connect with alumni and collect their career information. Current approaches for career analysis using LinkedIn data often require a labor-intensive manual process that is not scalable and sustainable. In this paper, we propose a system for the automated retrieval and classification of LinkedIn job titles for alumni career analysis. Our research prototype and experiment results show that the proposed system can effectively crawl LinkedIn profiles and classify job titles based on Standard Occupational Classification (SOC) job categories [1]. Our approach is the first of its kind in career analysis using LinkedIn and can be easily adopted by other universities or programs to develop career analysis systems for their alumni.

CCS CONCEPTS

Social and professional topics → Professional topics → Computing education → Computing education programs

KEYWORDS

Data analysis; LinkedIn; employment; career path; computing education.

ACM Reference format:

Lei Li, Svetlana Peltsverger, Jack Zheng, Linh Le, Michael Handlin. 2021. Retrieving and Classifying LinkedIn Job Titles for Alumni Career Analysis. In *Proceedings of ACM SIGITE 2021. October 6–9, 2021, SnowBird, UT, USA . ACM, New York, NY, USA, 6 pages.* https://doi.org/10.1145/3450329.3476858.

SIGITE '21, October 6–9, 2021, SnowBird, UT, USA

ACM ISBN 978-1-4503-8355-4/21/10...$15.00
https://doi.org/10.1145/3450329.3476858

1. Introduction

It has become increasingly important for higher education institutions to track their graduates' job placement and career development. The job placement rate is considered one of the important quality indicators for universities and colleges by the Department of Education. Accreditation agencies such as ABET ask the accredited programs to measure the program's educational outcomes that describe what graduates are expected to attain within a few years after graduation. These outcomes are usually more difficult to assess directly [2]. Moreover, universities can use graduates' career data to continuously improve programs, build alumni relationships, market, and conduct other broad range of activities. Many universities have relied on post-graduation surveys to collect their graduates' job information but often face challenges as the response rates of alumni surveys are usually low.

LinkedIn, with 756 million members in more than 200 countries and territories worldwide [3], is widely used by job seekers and recruiters due to the almost unlimited networking and career development opportunities. College students have increasingly used LinkedIn as a portfolio building and self-promotional tool to help them start their careers. Those students usually continue using LinkedIn to advance their careers after graduation. A LinkedIn profile usually contains an up-to-date and complete history of the user's education, jobs, and skills. Many universities utilize LinkedIn to build online communities at various levels and stay in touch with their students and alumni. More importantly, LinkedIn presents a gold mine for the universities to harvest often-hard-to-get alumni career information.

Researchers recognized the opportunity and started to use LinkedIn for alumni career analysis [4] [5]. However, collecting employment data from LinkedIn is not free of problems: 1) alumni LinkedIn profiles are often not readily available, and it is difficult and time-consuming to identify the correct alumnus profile; 2) the information retrieval from LinkedIn is often a labor-intensive manual process, and it is not sustainable in the long term; 3) the job titles in a LinkedIn profile are self-entered which makes the career analysis at a program level challenging.

To tackle those issues, this paper proposes an automated approach that can identify relevant LinkedIn profiles, retrieve information from the profile, and match job titles to a standard job classification defined by the US Bureau of Labor Statistics.

The rest of the paper is organized as follows: section two describes related studies in alumni career analysis using LinkedIn data, job title standardization, and classification; section three introduces the system architecture; section four presents the preliminary results, and section five concludes the paper and discusses limitations and future directions.

2. Related Works

2.1. LinkedIn Profile Identification and Information Retrieval

There are a few studies conducting alumni career analysis using LinkedIn data. Case et al. [4] collected job information from a closed LinkedIn alumni group and analyzed the career paths of those graduates. Case et al. [4] used a manual process in which alumni were searched individually by names, and they were invited to connect, then they join the closed LinkedIn group.

Li et al. [5] analyzed alumni's career trajectory, including entry positions and positions held in five, ten, and more than 10 years after graduation, by pulling alumni's job information from their LinkedIn profiles. In this study, the LinkedIn profile URLs were readily available as the students in the program were required to create LinkedIn profiles at the time of graduation. The researchers still had to retrieve career data from the Linked profiles manually.

While both Case et al. [4] and Li et al. [5]'s approaches demonstrated the potential and power of LinkedIn data, the manual process of identification and retrieval of LinkedIn profiles remain challenging.

2.2. Job Title Standardization & Classification

Another difficulty in using LinkedIn for career analysis is that LinkedIn job titles are self-entered by profile owners. Similar jobs might be named differently, and some job titles appear to be disorganized. For example, Bekkerman and Gavish [6] found almost 40,000 different ways users specified "Software Engineer." For another example, titles like "database administrator and designer" cover two or more job positions.

Li et al. [5] addressed the divergent LinkedIn job titles by mapping them into 11 researcher-defined job categories. While Li's approach worked in the scope of their research, it may not work for other computing-related programs in other institutions. In addition, mapping LinkedIn job titles to job categories was completed manually, which is not scalable when the dataset grows bigger.

Javed et al. [7] [8] adopted a standardized hierarchy of job titles commonly known as Standard Occupational Classification (SOC). The SOC system is used by federal statistical agencies to classify workers and jobs into occupational categories to collect, calculate, analyzing, or disseminating data [1]. There are several advantages of using a SOC system.

- Jobs are already organized into categories.
- Common occupation assessment tools are ready.
- It is easy to compare attributes across jobs and organizations or academic programs.

There are also some issues using standard job titles from SOC 2018.

- The standards are not flexible and specific enough to accommodate some unique situations.
- In the SOC system, categories are not exclusive to one other. Some jobs can be placed in several categories. Especially in IT, many positions take multiple roles and job functions.
- Due to changes in the job market, the standard needs to be frequently updated. It might not reflect recent trends and newly created jobs, especially in the Information Technology industry. Many of the titles in IT are too overarching, while some are too specific and focused.
- There is no hierarchy or other kinds of relationship among job titles to model advancement in career paths.

We believe that the benefits of the SOC standard job titles outweigh the potential issues. Some of the SOC's shortcomings can be overcome by imposing additional human coding. For example, Javed et al. [8] also introduced Carotene, a machine learning-based semi-supervised job title classification system for an online job recruitment company, CareerBuilder.com. They used Support Vector Machine (SVM) [9] for coarse level classification and K-Nearest-Neighbor (k-NN) [10] for fine level classification.

Given the promising results from Javed and associates' study, the SOC system and machine learning-based classification methods could be a good solution for standardizing and classifying LinkedIn job data.

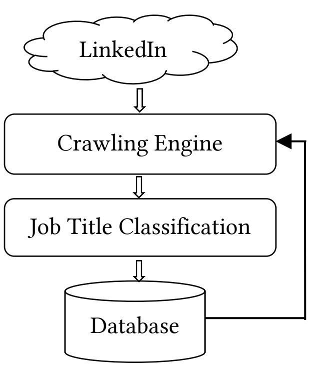

Figure 1. Automatic Job Title Crawling and Classification Architecture

3. System Architecture

To effectively perform career analysis using LinkedIn data, we propose a system that automatically crawls and classifies job titles from LinkedIn profiles. The system architecture is illustrated in Figure 1. The proposed system has three components:1) a crawling engine that retrieves job and other relevant information from a publicly available LinkedIn profile; 2) a job title classification system that maps the user self-entered job titles into standard job title defined by SOC; 3) a database system that stores the information for further data analytics.

3.1. Crawling Engine

The crawling engine retrieves job information such as the employer, position title, and the dates of employment from a LinkedIn profile. The system crawls using the URLs of LinkedIn profiles stored in the database. If URL is unknown, the system searches profiles through a search engine. Observing users' privacy, the system only crawls publicly available profiles, and the results will only be used for educational purposes.

All job titles pulled from LinkedIn are mapped to the standard job titles. Manual mapping is ideal for accuracy but impractical given the sheer number of data we are dealing with. The job title classification process is shown in Figure 2. We propose to use machine learning algorithms to classify the job titles from LinkedIn automatically. We tried several machine learning methods, which are detailed in the research prototype section.

3.2. Job Title Classification

We adopted the latest edition of SOC version [1], which lists 867 standard job titles and O*Net Online, a career exploration and analysis site sponsored by the US Department of Labor [11]. Given the context of this study, to identify IT-related job titles, we used the education crosswalks from O*NET [11]. We identified 20 IT-related job titles using Classification of Instructional Programs (CIP) code 11.xxxx. The 21 standard IT job titles are listed in Table 1.

Table 1. SOC IT Related Job Titles

Computer and Information Research Scientists	Computer Occupations, All Other
Computer and Information Systems Managers	Data Scientists
Computer Hardware Engineers	Database Administrators
Computer Network Architects	Database Architects
Computer Network Support Specialists	Information Security Analysts
Computer Programmers	Network and Computer Systems Administrators
Computer Science Teachers, Postsecondary	Software Developers
Computer Systems Analysts	Software Quality Assurance Analysts and Testers
Computer User Support Specialists	Web and Digital Interface Designers
Computer, Automated Teller, and Office Machine Repairers	Web Developers

We categorized non-IT-related jobs into job clusters as it is impractical to classify 847 job titles. We consolidated 15 job clusters defined O*NET into seven non-IT job clusters: Non-IT-Business, Non-IT-Education, Non-IT-Government, Non-IT-Health, Non-IT-STEM, Non-IT-Law & Public Safety, Non-IT-Arts. Therefore, there are 30 standard job titles in our classification system.

The SOC job titles we adopted don't capture the level of the jobs, and we believe that this information is important to analyze an alumni's career path. Moreover, after LinkedIn job titles are classified, they are evaluated on the technical level (non-technical, entry-level, intermediate/experienced, senior) and managerial level (non-managerial, first-line/team leads, mid-level, executives). For the non-IT-related job titles, only the managerial level is assessed. The classified job titles and levels are stored in the database.

The job title classification process is illustrated in Figure 2. The LinkedIn job titles are first searched for direct matches to the defined SOC job titles in the pre-classification phase. The remaining job titles are fed into the machine-learning based classification engine. Lastly, all classified job titles are assessed for levels by keyword matching through a computer program. For example, job titles contains keywords such as CIO, chief exectutive office, etc., will be categorized as senior level.

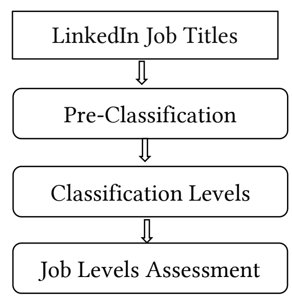

Figure 2. Job Title Classification Process

3.3. Alumni Database

The database is designed to store all the relevant information retrieved from LinkedIn profiles, standardized job titles and clusters, and job levels. The ERD diagram is shown in Figure 3.

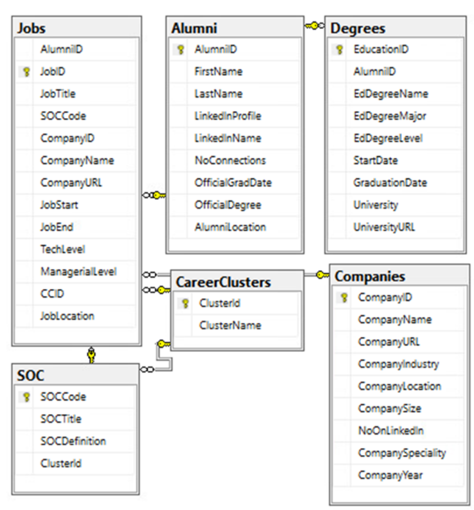

Figure 3. Database Structure

The StartDate and the GraduationDate are self-reported from the LinkedIn profile. The employment dates are set to January 1 if only the year of employment was reported. To distinguish between non-reported employment dates and current employment, 1/1/3000 is used to indicate the current employer.

4. Research Prototype and Preliminary Results

4.1. LinkedIn crawling

To make the proposed system scalable, we developed an automatic mechanism to collect career information from LinkedIn profiles. We intend to use the proposed system for building an alumni database for higher education institutions and assume that we either have the LinkedIn profile URLs or student names from the previous alumni database or graduation records.

Identification of LinkedIn profile using names. We used Python scripts and Bing API v7 to search for a combination of the alumni name and the university name. The first five results of the Bing search were checked for inclusion of the "linkedin.com/in/" in the URL to identify user profiles. For example, "query = linkedin+john+doe" generated five results, and only the last three profile URLs were used for further analysis.

1. John Doe | LinkedIn
   https://www.linkedin.com/company/john-doe
2. 8,800+ "John Doe" profiles | LinkedIn
   https://www.linkedin.com/pub/dir/John/Doe
3. John Doe - Corporate Executive - Baller Inc. | LinkedIn
   https://www.linkedin.com/in/john-doe-882117212
4. john doe - yes - no | LinkedIn
   https://www.linkedin.com/in/john-doe-a57060212
5. John Doe - Founder - Citrus Labs Limited. | LinkedIn
   https://ke.linkedin.com/in/john-doe-547199212

The LinkedIn profile URLs were then screened based on the user's educational background. Once the LinkedIn profile is identified, we use Python script with Selenium web driver and Parsel parser to crawl the profile and collect the location, number of connections, places, dates of employment, employers' LinkedIn profiles, education records, etc. LinkedIn limits how many profiles per hour can be accessed, and this task runs in the background with 2-5 minutes between requests. An essential consideration in this process is that the users enter the job titles on the LinkedIn profiles, and as a result, similar jobs might be named differently.

For this paper, we started with our alumni database, which contains LinkedIn profiles of graduates from an IT graduate program in a large public university in the southeast of the US. We also used recent graduates' names to identify their public LinkedIn profiles.

4.2. Job Title Classification.

The job title classification is divided into 4 phases: 1) data encoding; 2) training dataset development; 3) machine learning; 4) continuous improvement of the model.

Data encoding. The job titles retrieved from LinkedIn profiles are text data. We used the Universal-Sentence-Encoder developed by Google and widely used by the text mining community [12] to convert the texts into high dimensional vectors fed into the machine learning algorithms. For example, "application developer" is encoded as a vector with 512 values.

[0.0596 0.0356 … -0.0040 0.0604]

Training dataset development. Having a good training dataset is important for the performance of machine learning algorithms. The LinkedIn job titles are first manually mapped to the standard job titles by three faculty experts. The faculty experts first discussed the manual coding strategies. Then each expert mapped one-third of the job titles. After initial coding, the faculty experts met and cross-referenced their coding results to ensure the coding was done consistently among all three experts. 1300 LinkedIn job titles were manually coded.

Machine learning. We used three commonly used machine learning algorithms to classify the job titles retrieved from LinkedIn profiles: Random Forest (RF) [13], Neural Network (NN) [14], Support Vector Machine (SVM) [9]. As some job titles have low frequencies in our dataset, we also trained the models with a weighted class strategy available in RF and SVM.

The implementations of our chosen machine learning algorithms were adopted from scikit-learn.org, a Python library for machine learning [15].

Standard job titles with less than 25 matches were removed from the data as the machine learning algorithms generally need relatively populated classes in the dataset as prediction targets. After filtering the low-frequency classes, 1142 LinkedIn job titles mapped into 14 SOC 2018 standard job titles are used as a dataset for the machine learning algorithms. The dataset is further divided into two groups: training data (70%) and testing data (30%).

All machine learning algorithms are fine-tuned in the training data with the below settings and hyperparameters:

- RF: number of trees in {20, 50, 100}; minimum data points for split in {10, 20, 30}; and minimum data points in a leaf in {10, 20, 30}.
- SVM: using polynomial kernel with degree of two or three, or using Gaussian kernel with gamma in {0.001, 0.01, 0.1, 1, 10, 100, 1000}; and regularization term C in {0.01, 0.1, 1, 10, 100}.
- NN: using Rectified Linear Unit (ReLU) activation and SoftMax output. The networks have two or three layers. The number of neurons in each layer is in {n/2, n, 2n} where n is the input dimension (which is 512); and regularization term α in {0.01, 0.1, 1, 10, 100}.

We adopted commonly used evaluation metrics for machine learning: true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN). These measurements are then used in calculations for four values: accuracy, the rate of a TP or TN compared to the sum of all detections to gauge overall correctness; precision, which measures the accuracy of detecting TPs against FPs; recall, which determines a rate for how many FPs were reported compared to the total true threats; and F1-Score, which provides another rate for overall accuracy [16]. The formulas for each are presented as follows.

Accuracy = (TP + TN) / (TP + FP + FN + TN)
Precision = TP / (TP + FP)
Recall = TP / (TP + FN)
F1-score = 2 * Precision * Recall / (Precision + Recall)

We used accuracy and F1 values to measure the performance of the machine learning algorithm. Each algorithm was trained five times using the selected dataset with random seeds. The accuracy and F1 value of each model from each run and average values of 5 runs are shown in Tables 2 and 3, respectively. RF without weighted class is excluded from the report as its performance is considerably worse than other models.

Table 2. Machine Learning Models Performance

Runs	Metrics	RF (weighted)	NN	SVM	SVM (weighted)
#1	Accuracy	0.6676	0.7026	0.7055	0.6968
	F1 value	0.6654	0.6906	0.7312	0.7206
#2	Accuracy	0.6356	0.7026	0.7259	0.7347
	F1 value	0.6792	0.7011	0.7293	0.7390
#3	Accuracy	0.6939	0.7201	0.758	0.7638
	F1 value	0.7180	0.7344	0.7815	0.7908
#4	Accuracy	0.6676	0.6980	0.7026	0.7143
	F1 value	0.6858	0.6954	0.7029	0.7175
#5	Accuracy	0.7114	0.7318	0.7726	0.7726
	F1 value	0.7284	0.7671	0.7908	0.7918

Table 3. Average Machine Learning Models Performance

Algorithms	Accuracy	F1 Value
RF (weighted)	0.6752	0.6954
NN	0.7110	0.7177
SVM	0.7329	0.7471
SVM (weighted)	0.7364	0.7519

Among the four algorithms we used, SVM weighted generated the best classification models in 4 of the five runs. And SVM weighted also produced the best performance in both accuracy and F1 value when classifying the testing dataset. While a relatively small dataset was used to training the machine learning algorithms, the accuracy (0.7364) and F1 value (0.7519) of SVM weighted are fairly good and promising. With more training data becomes available, we expect the performance of machine learning can be further improved.

Continuous improvement of classification models. Machine learning algorithms often rely on large training data for better performance. One example application of the proposed system is to keep track of the alumni career information for academic units of a higher education institution. At the end of each academic semester, the LinkedIn profile will be pulled for graduating students. For the students already in the alumni database, updates will be run on their LinkedIn profiles. The classification models will be retrained with added data points. We believe that the classification models will reach a stable status after several rounds of retraining and be able to automatically classify the incoming job titles.

CONCLUSION AND DISCUSSION

Being able to track the job placement rates and career development is important for academic programs. In this paper, we proposed a system for the automated retrieval and classification of LinkedIn job titles which is the foundation for scalable and sustainable alumni career analysis. A system architecture was presented, and a research prototype was built. The experiment results showed that the proposed system could effectively crawl LinkedIn profiles and classify the job titles based on defined standard job categories.

Based on our knowledge, this work is the first attempt for automated retrieval and classification of LinkedIn job data for alumni career analysis in higher education. Other universities or programs can easily adopt our approach to develop alumni career analysis systems for their programs.

There are some legality and ethics implications of scraping LinkedIn website. This project is intended for educational uses only. We only retrieved data from publicly available LinkedIn profiles and will only report results using aggregate data without any identifiable information. Moreover, we have requested API access to extended profiles using LinkedIn r_fullprofile and plan to use an authorization code flow to request permission from LinkedIn members to access their account data.

There are several limitations to this study. First, our method for automatic retrieval of LinkedIn profiles is web page HTML scrapping. We need to modify our code accordingly if LinkedIn changes its webpage structure. LinkedIn also limits the number of searches that can be performed per day. Therefore, it may take longer to crawl the LinkedIn for large universities.

Secondly, it's often difficult to locate the right profile in LinkedIn by names only due to the following reasons.

- Using preferred first names instead of official names (especially by international students).
- Legal name changes.
- Multiple profile matches for common names (employees, students who graduated in different years, or other degrees).
- Private profiles.
- Bing API may generate different search results in different query attempts.

The academic programs could encourage their graduating students to share their LinkedIn profiles as part of an exit interview or a capstone course to mitigate such challenges.

Thirdly, this study's machine learning algorithms need to be fine-tuned with more training sets. While continuous improvement is built in our proposed system, it takes time for the classification engine to reach a mature and stable stage. Moreover, job titles from LinkedIn have some limitations for analysis. For example, based on job responsibilities, a "Solution Architect" can be a "Data Architect" or a "Computer Engineer" (indeed.com). Burning Glass Technologies, a job market analytics software company, classifies job ad titles based on the job requirements. Our classification can also be improved by linking a job title to a job ad from Burning Glass. For instance, an "Application Developer at Texas Engineering Extension Service in College Station, TX" job title combined with employment dates from a LinkedIn profile can be classified as a web developer based on a job ad from Burning Glass.

We plan to expand our research in the following directions.

1) This study is research-in-progress. The accuracy of machine learning based job title classification is about 73%. Manual review and verification are still needed. We need to refine our research prototype and improve the effectiveness of the classification engine by creating more training data through manual coding.

2) Expand the proposed system to collect more career-related information. In addition to the job titles, it is also important to understand the skills or even salary related to a job title [17]. We plan to use an O*NET API and Pandas [18], a flexible and powerful data analysis Python library.

3) We currently focus on the LinkedIn data retrieval and job title classification. The next step is to store the processed data in the database for further analyses. The job placement rate of graduates can be easily calculated. We can examine the correlation between alumni' academic performance at school and the career path they choose after graduation. We can also correlate skills and job titles. It will also be interesting to investigate the career path of the alumni from undergraduate and graduate programs in the same discipline as well as career paths in related computing disciplines. For example, we can compare the career development of graduates from an undergraduate and a graduate IT program. We also can collaborate with colleagues from other disciplines and find out how the career path differs for graduates from Information Technology, Computer Science, Information Systems, and Software Engineer.

4) We also need to find a better way to define a person's career path through the relationship and progression of job titles. Traditional career path includes laddered (linear) progression and lattice pathways. IT career path is very dynamic and complicated, and we have yet seen any formal definition and categorization of IT career paths in recent studies. At this time, career paths are simply modeled as seniority levels and IT/Non-IT (management) transitions. A career path should also include progression/movement among job functions (development, support, research, etc.), domains (data, system, security, etc.), skill requirement/growth, business sectors/industries. It also needs to acknowledge additional training and education in emerging technologies.

5) Some job titles from LinkedIn profiles contain job descriptions or responbilities. That information can be analyzed and cross-referenced with the LinkedIn user' self-reported skills. The analysis results can help academic programs to determine whether their graduates obtained certain skills as required by accreditation agencies such as ABET.

REFERENCES

[1] US Bureau of Labor Statistics, 2018, 2018 Standard Occupational Classification System, https://www.bls.gov/soc/2018/home.htm, last retrieved on May 12, 2021.

[2] ABET, https://www.abet.org, last retrieved on May 12, 2021.

[3] LinkedIn, 2021, https://about.linkedin.com/ , last retrieved on may 12th , 2021.

[4] Case, T., Han, H., and Rimes, E., 2016, Career Paths of Computing Program Graduates: A LinkedIn Analysis, SAIS 2016 Proceedings. 21. http://aisel.aisnet.org/sais2016/21

[5] Li, L., Zheng, G., Peltsverger, S. and Zhang, C., 2016, Career Trajectory Analysis of Information Technology Alumni: A LinkedIn Perspective"." In Proceedings of the 17th Annual Conference on Information Technology Education, 2–6. Boston Massachusetts USA: ACM. https://doi.org/10.1145/2978192.2978221.

[6] Bekkerman, R., and Gavish, M., 2011, "High-Precision Phrase-Based Document Classification on a Modern Scale." In Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD' '11, 231. San Diego, California, USA: ACM Press. https://doi.org/10.1145/2020408.2020449.

[7] Javed, F., McNair, M., Jacob, F., and Zhao, M., 2014. Towards a job title classification system. In WSCBD 2014: Webscale Classification: Classifying Big Data from the Web, WSDM Workshop O*Net online, 2021, Occupation Search, https://www.onetonline.org/ , last retrieved on May 12, 2021.

[8] Javed, F., Luo, Q., McNair, M., Jacob, F., Zhao M., and Kang, T. S., 2015, "Carotene: A Job Title Classification System for the Online Recruitment Domain," 2015 IEEE First International Conference on Big Data Computing Service and Applications, pp. 286-293, doi: 10.1109/BigDataService.2015.61.

[9] Suykens, J. A., & Vandewalle, J. (1999). Least squares support vector machine classifiers. Neural processing letters, 9(3), 293-300.

[10] Guo, G. D., et al. 2006. Using kNN model for automatic text categorization. Soft Computing, 10(5): 423-430.

[11] O*NET Online, 2021, https://www.onetonline.org, last retrieved on May 30, 2021.

[12] TenserFlow, https://tfhub.dev/google/universal-sentence-encoder, last retrieved on May 12, 2021.

[13] Pal, M. (2005). Random forest classifier for remote sensing classification. International journal of remote sensing, 26(1), 217-222.

[14] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. nature, 521(7553), 436-444.

[15] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2021, Scikit-learn: Machine Learning in Python, Journal of Machine Learning Research, 12(85):2825–2830.

[16] Masum, M., and Shahriar, H. 2020. TL-NID: Deep Neural Network with Transfer Learning for Network Intrusion Detection, in 2020 15th International Conference for Internet Technology and Secured Transactions (ICITST), London, United Kingdom: IEEE, December 8, pp. 1–7.

[17] Zheng, J. & Zhang, C., & Li, L., 2015. Practicing and Evaluating Soft Skills in IT Capstone Projects. SIGITE '15: Proceedings of the 16th Annual Conference on Information Technology Education September 2015 Pages 109–113https://doi.org/10.1145/2808006.280804110.1145

[18] Pandas https://github.com/pandas-dev/pandas, , last retrieved on May 12.