**USING NATURAL LANGUAGE PROCESSING TO IMPROVE CAREER PROSPECTS FOR COMPUTER STUDENTS**

**BY**

**JOSHUA KIPRONO**

**A Thesis Submitted in Partial Fulfillment of the Requirements**

**For the Degree of Bachelor of Science in The**

**Department of Mathematics and Computer Science**

**Claflin University**

**Orangeburg, South Carolina**

**MARCH 2023**

|  |  |  |
| --- | --- | --- |
| **DR. SHRIKANT PAWAR** |  | **DR. WEI WAN** |
| **Thesis Advisor** |  | **Committee Member** |
|  |  |  |
| **DR. ABDELRAHMAN DESOKY** |  | **DR. RAMAIER SRIRAM** |
| **Committee Member** |  | **Academic Advisor** |

|  |  |  |  |
| --- | --- | --- | --- |
| **Received by:** |  | **Date:** |  |
| **Dr. Karina Liles, Chairperson, Department of Mathematics and Computer Science** | | | |
| **Received by:** |  | **Date:** |  |
| **Dr. Verlie Tisdale, Acting Dean, School of Natural Sciences and Mathematics** | | | |
| **Received by:** |  | **Date:** |  |
| **Dr. Verlie Tisdale, Interim Vice Provost for Academic Affairs** | | | |

**ACKNOWLEDGEMENTS**

I would like to extend my sincere gratitude to Dr. Shrikant Pawar who served as my thesis advisor, for his continuous support, direction, and mentorship during my thesis project. His priceless comments, helpful critique, and professional guidance have greatly influenced my study and assisted me in reaching my thesis objectives.

I also want to thank my thesis committee, for their constructive criticism, perceptive remarks, and helpful recommendations as my thesis was being developed. I am extremely appreciative of their support and contributions, which have been essential to the success of my thesis.

In addition, Dr. Sriram, who has been a continual source of inspiration and encouragement for me throughout my academic career, deserves recognition for the advice and help he has given. His advice and support have been crucial in assisting me in overcoming hurdles and achieving my academic objectives.

Finally, I want to express my gratitude to my family and friends for their constant support and inspiration during my academic career. They have provided me with inspiration and motivation, and I am incredibly appreciative of their presence in my life.

**THESIS STATEMENT**

The increasing demand for computer science graduates in today's job market has created a fiercely competitive landscape, making it difficult for graduates to effectively communicate their skills and experience to potential employers through traditional resumes. According to a recent survey by the National Association of Colleges and Employers (2020), only 24% of employers consider GPA and coursework to be the most important factors when evaluating candidates, while 77% rank relevant work experience as the top priority. This research project aims to bridge the gap between job seekers and employers by leveraging Natural Language Processing (NLP) techniques to match the skills and qualifications listed on resumes with relevant job opportunities while providing career advisors with an efficient and effective way to manage and review resumes.

Previous research has shown that Natural Language Processing (NLP) can effectively extract and classify information from resumes and job postings (Lin & Huang, 2019), and Machine Learning algorithms can accurately match skills and job requirements (Garg et al., 2018). However, there is still a lack of research that focuses on using Natural Language Processing (NLP) to match the skills and qualifications of computer science graduates with relevant job opportunities. By building upon previous research findings primarily focused on Natural Language Processing (NLP) and Machine Learning algorithms in the last 5 years, the project aims to address the meaningful gap in the current research literature and demonstrate the potential of Natural Language Processing (NLP) in simplifying the job search process for computer science graduates and streamlining the resume review process for career advisors.

Bottom of Form

# 

**ABSTRACT**

As technology advances, the job market for Computer Science graduates has become highly competitive, making it essential for students to have the technical skills and the ability to communicate their qualifications and experience to potential employers through their resumes. Career advisors face the challenge of managing and tracking students' resumes to help them find the best job opportunities. To address this, this research project has been developed to assist students in identifying job opportunities that align with their skills and experience while simplifying the process for career advisors.

The project uses Natural Language Processing (NLP) techniques, starting with entity recognition on students' resumes to identify and classify named entities such as skills, qualifications, and experience, and organize them into structured data. Machine Learning algorithms are used to match the skills extracted from the resume with the requirements listed in the job postings, providing a ranked list of job recommendations to the job seeker.

The prototype was tested on resumes reviewed by the Claflin University Career Services department, and it proved successful in identifying career paths and job opportunities. A user study conducted by the Claflin University Computer Department found that the project improved the career matching process, making it more efficient and presenting job opportunities that were better aligned with the skills and qualifications outlined on the resumes.

The project offers a convenient and efficient way for students to identify job opportunities that align with their skills and experience and streamlines the resume review process for career advisors. The project demonstrates the potential of Natural Language Processing (NLP) to facilitate the job search process and improve the matching of skills and job requirements.

**KEYWORDS AND ABBREVIATIONS**

AI – Artificial Intelligence

ATS – Applicant Tracking System

NLP – Natural Language Processing

NLTK – Natural Language Toolkit

**LIST OF TABLES**

[Table 1. Observed skills compared to retrieved Skills on test resume 15](#Table_1)

[Table 2. Observed skills compared to retrieved Skills on test resume 15](#Table_2)

[Table 3. Observed skills compared to retrieved Skills on test resume 16](#Table_3)

[Table 4. Observed skills compared to retrieved Skills on test resume 17](#Table_4)

[Table 5. Observed skills compared to retrieved Skills on test resume 18](#Table_5)

[Table 6. Observed skills compared to retrieved Skills on test resume 19](#Table_6)

[Table 7. Observed skills compared to retrieved Skills on test resume 20](#Table_7)

[Table 8. Observed skills compared to retrieved Skills on test resume 21](#Table_8)

[Table 9. Observed skills compared to retrieved Skills on test resume 22](#Table_9)

[Table 10. Observed skills compared to retrieved Skills on test resume 23](#Table_10)

**LIST OF FIGURES**

Figure 1. *The process of extracting meaningful text from a text corpus* 9

[Figure 2. Pie Chart showing *top 5 career interest matches* 24](#Image_2)

[Figure 3. *Pie Chart showing satisfaction with search results* 25](#Image_3)

[Figure 4. *Pie chart showing satisfaction with search results* 26](#Image_4)

[Figure 5. *Pie chart showing the likelihood of using the application.* 27](#Image_5)

[Figure 6. *Pie chart showing a likelihood of users sharing results of the analysis* 28](#Image_6)

[Figure 7. *bar chart showing skill extraction accuracy for test resume****s.*** 28](#Image_7)

[Figure 8. *bar chart showing the percentage of extra skills extracted from test resumes****.*** 28](#Image_7)

**TABLE OF CONTENTS**

[**ACKNOWLEDGEMENTS** i](#_Toc112781807)

[**THESIS STATEMENT** ii](#_Toc112781808)

[**ABSTRACT** iii](#_Toc112781809)

[**KEYWORDS AND ABBREVIATIONS** iv](#_Toc112781810)

[**LIST OF TABLES** v](#_Toc112781811)

[**LIST OF FIGURES** vi](#_Toc112781812)

[**TABLE OF CONTENTS** vii](#_Toc112781813)

[**INTRODUCTION** 1](#_Toc112781814)

[Background of Study 1](#_Toc112781815)

[Purpose of Study 1](#_Toc112781816)

[Research Question(s) and Hypotheses 2](#_Toc112781817)

[Summary 2](#_Toc112781818)

[**LITERATURE REVIEW** 3](#_Toc112781819)

[Summary and Conclusions 3](#_Toc112781820)

[**METHODOLOGY** 4](#_Toc112781821)

[Population 4](#_Toc112781822)

[Sampling and Sampling Procedures 4](#_Toc112781823)

[Data Collection Procedures 4](#_Toc112781824)

[Instrumentation (if applicable) 5](#_Toc112781825)

[Data Analysis Plan 5](#_Toc112781826)

[Limitations 6](#_Toc112781827)

[Ethical Procedures 6](#_Toc112781828)

[Summary 7](#_Toc112781829)

[**RESULTS** 8](#_Toc112781830)

[Data Collection 8](#_Toc112781831)

[Study Results in 8](#_Toc112781832)

[Summary 9](#_Toc112781833)

[**DISCUSSION AND CONCLUSIONS** 10](#_Toc112781834)

[Summary of the Findings 10](#_Toc112781835)

[Recommendations 10](#_Toc112781836)

[Conclusions 11](#_Toc112781837)

[**REFERENCES** 12](#_Toc112781838)

[**Appendix A: Title of Appendix** 13](#_Toc112781839)

# **INTRODUCTION**

# **Background of the Study**

The rapid evolution of technology has led to the emergence of numerous new and specialized technical fields, creating an increasingly competitive job market for Computer Science college graduates. As a result, it has become imperative for students to not only possess the necessary technical skills but also to effectively communicate their qualifications and experience to potential employers through their resumes.

Through a review of relevant literature and data analysis, this paper aims to provide insight into the current state of the job market and the specific technical fields that are in high demand. Additionally, this paper will also explore the use of Natural Language Processing (NLP) techniques to analyze resumes and job descriptions, identify patterns and keywords, and make suggestions for improving resumes. By utilizing Natural Language Processing (NLP) techniques, students can stay up to date with the current trends in technical skills and tailor their resumes to match the requirements of the job market.

In conclusion, this paper aims to make a significant contribution to the ongoing conversation about the employability of college graduates in the rapidly evolving technology landscape. It is hoped that the findings of this study will inform future career development strategies and educational practices, ultimately leading to the better preparation of college graduates for the workforce by highlighting the importance of Natural Language Processing and Artificial Intelligence (AI) in general as a solution to help students adapt to the evolving trends in technical skills.

**Purpose of the study**

The purpose of this research paper is to create a Natural Language Processing (NLP) platform and then use the platform to investigate Natural Language Processing (NLP) techniques. The techniques involve parsing resumes and extracting skills from the documents, matching the extracted skills with job listings to identify job opportunities that align with the job seeker's skills, and a quantitative survey of its impact and accuracy. This is aimed at effectively showcasing a student’s qualifications and experience in these specialized technical fields to adapt to these evolving trends and increase their employability.

Research Question and Hypotheses

The following are the research questions posed:

1. How accurately can Natural Language Processing (NLP) techniques extract skills from resumes?
2. How effectively can the extracted skills be matched with job listings to identify job opportunities?
3. How does the use of Natural Language Processing (NLP) to match skills and job requirements impact the job search process for job seekers?

The hypothesis of the research is the following:

1. The Natural Language Processing (NLP) techniques used in the project can accurately extract skills from resumes.
2. The extracted skills will effectively match job listings to identify job opportunities that align with the job seeker's skills and experience.
3. The use of Natural Language Processing (NLP) to match skills and job requirements will improve the job search process for job seekers and the hiring process for employers, reducing the time and effort required to find the right person for the job.

**LITERATURE REVIEW**

This literature review examines the job market outlook for computer science graduates, which has been positively impacted by the increasing prevalence of technology. However, the high level of competition among graduates, particularly in a job market where Applicant Tracking Systems (ATS) are widely used by recruiters, creates challenges for job seekers. The ATS filters applications for specific keywords and qualifications, and if an applicant's resume lacks those, they may not be considered, even if they possess the necessary skills for the job. To address these issues, the study suggests that Natural Language Processing (NLP), a branch of AI, can play a crucial role in enhancing graduates' employability by analyzing resumes, and job descriptions, and identifying patterns and keywords. Overall, the study suggests that AI and Natural Language Processing (NLP) can play a crucial role in helping students adapt to the evolving trends in technical skills and the job market, ultimately leading to better career prospects for computer science graduates.

## **Outlook of the job market on computer science college graduates**

The increasing prevalence of technology has resulted in a heightened demand for graduates with degrees in computer science. This is due to the projected growth of employment opportunities for computer and information research scientists, which is anticipated to expand by 21% from 2021 to 2031 [1], significantly outpacing the average growth rate across all occupations. According to the US Bureau of Labor Statistics [2] there are over 20 different occupations within the field of computer science. This diversity in career options for computer science graduates and the increase in job demand for this field may appear to be a positive development on the surface, but upon further examination, this situation may also have some underlying consequences. One of the primary concerns is the high level of competition for jobs among computer science graduates. More companies no longer need degrees for positions and the number of graduates entering the job market has steeply risen over the years [3]. The unemployment rate for US computer science stands at 7.8% [4]. This relatively high rate means that the number of available positions, despite needing workers, is not keeping pace with the influx of job seekers. This creates a sharp competition among college graduates for available jobs.

## **The impact of technology (Applicant Tracking Systems) on hiring trends.**

Aside from the increasing competition, the majority of Fortune 500 companies are using Applicant Tracking Systems. Applicant Tracking Systems are a type of software used by companies to assist with human resources, recruitment, and hiring processes. While the use of (Application Tracking System) ATS significantly boosts performance for recruiters and companies [5], they may offer a worse experience for graduates seeking jobs in their fields. A recruiter has the option to reduce the number of applicants by filtering for keywords rather than looking through the hundreds of applications that are submitted for a position [6]. They could be job descriptions, technical skills, or qualifications. Even if a candidate is exceptionally qualified for the position, they might not be noticed if their resume doesn't contain the correct terms. In a competitive job market and combined use of applicant tracking systems, it is more important for graduates to be more specific in their resumes and tailor them with keywords that would pass through such tracking systems.

## **The role of Artificial Intelligence (NLP) in improving career prospects for students.**

Natural language processing is a branch of artificial intelligence that focuses on the interaction between computers and human languages, in both written and spoken forms. The goal of Natural Language Processing (NLP) is to enable computers to understand, interpret, and generate human language in a way that is both natural and useful [7]. By analyzing vast volumes of data and seeing patterns and trends in job choices, Natural Language Processing (NLP) can significantly improve career matching for students. Based on each student's unique strengths, interests, and aspirations, this information can be used to offer them individualized career advice and recommendations. Additionally, by examining resumes, cover letters, and job descriptions to determine the most qualified individuals for a position, Artificial Intelligence (AI) and Natural Language Processing (NLP) can be utilized to automate the process of matching students with employment possibilities. This can reduce the amount of time and resources needed for both students and businesses to locate a good match and raise those odds.

# **METHODOLOGY**

## **Introduction**

The purpose of this study is to examine the effectiveness of Natural Language Processing (NLP) as a technique to improve students' career prospects. Students will be able to submit their resumes for examination using a Natural Language Processing (NLP) based platform. The resumes will be assessed for the use of clichés and passive verbs, as well as for their ability to accurately identify the most likely career path given the skills listed. The software will also offer a ranking of relevant career possibilities that match the skills listed on the résumé. The ultimate purpose of this study is to pinpoint areas on student resumes that need work and to gauge how Natural Language Processing (NLP) based resume analysis affects prospects for professional progression.

**Population and Sample**

The target population is the Computer Science and Computer Engineering students at Claflin University, comprising about 101 students as per the Claflin University 2021-2022 fact book.

**Data Collection Procedures**

The data collection process in this study was designed to gather valuable insights into the effectiveness and accuracy of the Natural Language Processing (NLP) platform in improving career prospects for computer students. To this end, two phases of data collection were employed.

The first phase involved a visual analysis test with a Career Advisor member of Claflin University. This test was conducted to assess the accuracy and performance of the platform and make any necessary adjustments before the second phase. For the test, 10 anonymized resumes were selected from various sources, including student alumni, LinkedIn, and online sources. The Career Advisor was asked to analyze four aspects of the resumes, including their general organization, use of action words, description of skills, and skills that stand out. The results of this test were then compared to the results generated by the Natural Language Processing (NLP) platform to gauge its accuracy in terms of the action words identified, skills, and likely career paths.

The second phase involved a survey of students who had utilized the platform. Participants were recruited through personal contacts and referrals, and efforts were made to reach a diverse range of participants from different classifications. The study employed Google Form surveys to collect data, and participants were asked to complete a survey of 16 questions after they had used the platform and observed its operation. The online nature of the study was considered, and participants were provided with a secure and private link to access the online platform where the survey was hosted. Emphasis was placed on the privacy of data, as resumes often contain personally identifiable information. Participants were encouraged to answer the questions as honestly and accurately as possible, to the best of their understanding and knowledge.

The questions in the survey were divided into two parts: Part 1 focused on questions about the student, and Part 2 focused on questions about the Natural Language Processing (NLP) platform. The data collected from the surveys were analyzed to gain a better understanding of the platform's performance and its impact on the career prospects of computer students.

In conclusion, the data collection process was designed to gather meaningful and relevant information about the Natural Language Processing (NLP) platform and its impact on career prospects for computer students. The visual analysis test and the student survey were carefully planned and executed to ensure the validity and reliability of the data collected. This thesis shows how Natural Language Processing (NLP) can be used to improve career prospects for students. This involves skill analysis in resumes and ranked recommendations of career prospects.

To extract skills from the resume, the text is extracted from the uploaded resume. Due to the many text versions, this is achieved using the Python library, *PyPDF2*, an open-source pure-python PDF library capable of splitting, merging, cropping, and transforming the pages of PDF files. The corpus of text that is extracted is messy and contains a lot of noise. Pre-processing of the data is needed to make the text more suitable for Natural Language Processing (NLP) models that will be extracting important information. This can lead to better performance and more accurate results. The first step is data cleaning steps such as removing special characters, numbers, and stop-words, and lowercasing all text can help to clean the data and make it more consistent.

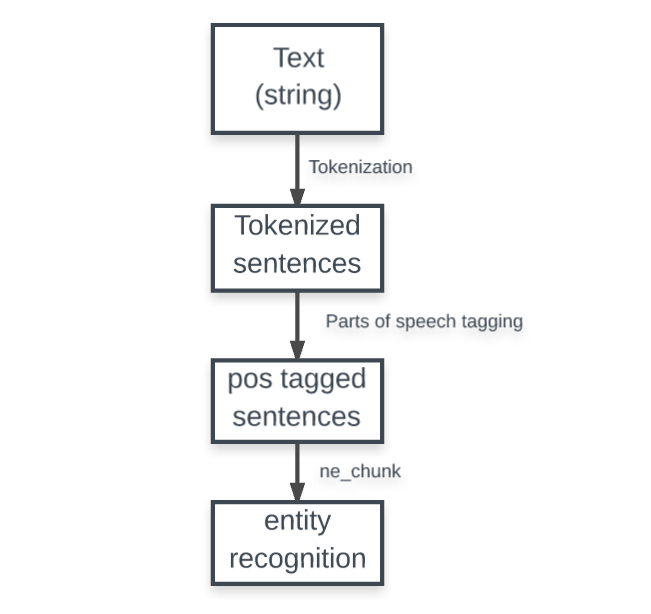
The second step in the process involves text normalization, which is a preprocessing step to convert unstructured text input into a more standardized and usable format. It aims to improve the text data's suitability for Natural Language Processing (NLP) tasks. The sub-steps involved in normalizing data include removing stop words and special characters, lowercasing, tokenization, and stemming.

Stop words are typically words such as "the", “and” "is", "our ", "a", and "an." These words appear frequently in text data but do not contribute much to the meaning of the text and are therefore considered to be noise that can be removed to improve the performance of Natural Language Processing (NLP) models. The process of removing stop words typically involves identifying a list of stop words and then removing them from the text data. To solve this problem, I used a library, *NLTK*, in python and downloaded a comprehensive list of English stop words.

Tokenization involves dividing a text document into tokens. Words, phrases, sentences, and even characters can be used as tokens. Tokenization facilitates the transformation of unstructured text input into a more organized format. In my problem, I tokenized my texts into words as finding skills involves singular words.

To identify action words in the resume, I used Parts of Speech (POS) tagging to assign a Parts of Speech (POS) label to each word in a text and find words labeled as verbs. Since there is no definitive method of finding the type of verb, I maintained a count of every verb found in the resume and if there are less than five verbs, that would designate the resume as a “passive”, a resume which is less compelling to potential employers.

To identify the most used words, I maintained a manually annotated bank of words and compared them against the tokenized words found in the resume. These words were then flagged as a potentially cliched words.



***Figure 1: The process of extracting meaningful text from a text corpus.***

To identify skills from unstructured data, entity recognition seemed like the ideal way to approach the problem. Entities in Natural Language Processing are specific words that represent real-world objects or concepts. However, the approach would be difficult to implement as skills can vary wildly from one person to another. While software skills may be common for this use case, the tools, and techniques used may differ i.e. JAVA and AutoCAD. These kinds of skills would not be identified as skill entities with the existing models. To solve this problem, I used dynamic analysis of existing skills. The idea here was to run every processed text against an existing database of open-sourced and verified skills. The database contains more than 25,000 skills, then uses fuzzy-matching and diff-algorithms to find closely matching words to find relevant skills.

Due to the variance in skills, I did a comparison on two levels: direct comparison and diffing to find closely matching words. For the first level, I did a direct string comparison, for example, some skills may be equal on the surface for instance if the word *Python* (programming language) occurs in both the resume and skill database. For the second level comparison, I did a diffing comparison to find the top five closest words. To achieve this, I used a diffing library called *difflib.*

The function *get\_close\_matches(word, possibilities, n, cutoff)*  from *difflib* accepts four parameters:

* word - the word to find close matches for in our list.
* possibilities - the list in which to search for close matches of a word.
* n (optional) - the maximum number of close matches to return. Must be > 0. The default is 3.
* cutoff (optional) - a float in the range [0, 1] that a possibility must score to be considered like a word. 0 is very lenient, and 1 is very strict. The default is 0.6.

Running tests showed that using a cutoff value of 0.5 and an *n* value of 5 was ideal for identifying words that could be considered skills. After identifying the skills, I iterated over the row of skills to find the most occurring professional careers that use the said skills. The top professional area was then passed into a function that retrieved data from the website Glassdoor to find jobs or internships related to the top professional area extracted from the platform.

**Limitations**

One limitation of the present study was the insufficient sample size. The research design originally intended to recruit a sample of 30 participants, however, only 16 individuals were successfully recruited to participate in the study due to a variety of factors, including low levels of participant interest, time constraints, and difficulties in identifying participants who met the established inclusion criteria. This limitation has implications for the generalizability of the findings, as a larger sample size would have facilitated the inclusion of a more diverse range of participants and a more representative sample of the target population. Furthermore, smaller sample sizes are associated with increased levels of variability in the results, which increases the likelihood of both false positive and false negative findings. Some of the biases likely encountered in the study include the following:

1. Selection bias: Since the participants were found through personal connections and recommendations, the sample size may not accurately reflect the intended population. This can result in a skewed depiction of the findings.
2. Response bias: Due to social desirability bias or apprehension about being assessed, respondents might not provide as truthful of an answer as possible to the survey questions.
3. Inaccurate results due to observer bias: The Career Advisor team member that carried out the visual analysis during the initial phase might have brought their prejudices and interpretations to the task.
4. Systematic mistakes: The data collection tool (Google Form surveys) may include flaws or problems that make it difficult to measure some factors effectively.

**Ethical Procedures**

In conducting this research, the highest priority was placed on ensuring the protection of participant privacy and confidentiality. To prevent the leak of personal data, several ethical procedures were implemented. First, all participants provided informed consent before participating in the study. This informed consent process included a detailed explanation of the study's purpose, procedures, and potential risks, as well as information on the participants' rights and how their data would be collected, stored, and used. Second, all data collected were de-identified, meaning that any information that could potentially identify the participants, such as their names and contact information, was removed. Third, the Natural Language Processing (NLP) platform was built to be serverless, meaning that no personal data from any resume could be collected or stored anywhere.

# **RESULTS**

The purpose of this project is to investigate how Natural Language Processing (NLP) techniques can be used to parse resumes and extract skills from the document and to match the extracted skills with job listings to identify job opportunities that align with the job seeker's skills and experience. The following were the research questions posed:

1. How accurately can Natural Language Processing (NLP) techniques extract skills from resumes?
2. How effectively can the extracted skills be matched with job listings to identify job opportunities?
3. How does the use of Natural Language Processing (NLP) to match skills and job requirements impact the job search process for job seekers?

The hypothesis of the research is the following:

1. The NLP techniques used in the project can accurately extract skills from resumes.
2. The extracted skills will effectively match job listings to identify job opportunities that align with the job seeker's skills and experience.
3. The use of NLP to match skills and job requirements will improve the job search process for job seekers and the hiring process for employers, reducing the time and effort required to find the right person for the job.

Bottom of Form

Data Collection

Data Collection was done in two phases. The first phase involved visual analysis with a Career Advisor member of Claflin University to test the application and the second was a survey done among students. The present study employed Google Form surveys to collect data. Participants were asked to complete a survey of 16 questions after they had finished using the tool and observed its operation. To accommodate the online nature of the study, participants were provided with a link to access the online platform where the survey was hosted. It was emphasized that the participants should answer the questions as honestly as possible and to the best of their understanding and knowledge.

The questions were divided into two parts:

Part 1: Questions about the student

Part 2: Questions about the Natural Language Processing (NLP) platform

Study Results

The platform was tested with 10 sample resumes of students in computer science and computer engineering. The test involved visual and analytical analysis with Douglas Barnes, a faculty member in the Career Advising office. The test study aimed to determine the following use of action words and a description of skill and skills that stand out.

**Test 1**

**Below is the original resume used in the case study:** Attached as Appendix A

**Expected Career Area**: Software Engineering.

**Top Career Area from the platform (100% match):** Software Development, Computer Programming, Computer and Information Research Scientist, Information Security Analyst, Computer Systems Analyst, Architectural Engineering

|  |  |
| --- | --- |
| **Observed Resume Skills** | **Resume Skills retrieved by the NLP platform** |
| C++, Go, Python, HTML, CSS, JavaScript, SQL, TensorFlow, Django, jQuery, NumPy, OpenCV, Research | Programming, Research, TensorFlow, JavaScript, Technical, Servers, Engineering, HTML, Radar, Ui, NumPy, Mathematics, Os, OpenCV, CSS, investigation, Requests, C++, AWS, Django, Visual, Robot, Python, Video, Updates |

***Table 1: Observed Resume Skills compared to retrieved skills on Test Resume 1***

**Test 2**

Below is the original resume used in the case study: Attached as Appendix B

**Expected Career Area:** Software Engineering, Leadership

Top Career Area from platform (100% match): Marketing Management, Architectural and Engineering Management, Management Analysis, Computer and Information Research, Computer Systems Analysis, Information Security Analysis

***Table 2: Observed Resume Skills compared to retrieved skills on Test Resume 2***

**Career Office Advice:** Resume needs better formatting.

**Platform Advice**: Great use of active verbs, the occurrence of commonly overused words in the resume

**Test 3**

Below is the original resume used in the case study: Link: Attached as Appendix C

**Expected Career Area:** Data Analytics, Software Development

**Top Career Area from the platform (100% match):** Architectural and Engineering Management, Management Analysis,Computer and Information Research, Computer Systems Analysis, Computer Programming, Web Development

|  |  |
| --- | --- |
| **Observed Resume Skills** | **Resume Skills retrieved by the NLP platform** |
| PostgreSQL, Java, Coding, NumPy, matplotlib, Streamlit, Pandas, C, Seaborn, API, Scikit-learn, Machine Learning, Docker, Jira, YAML, Python, Calculus, Splunk, Research, Machine learning, | Programming, Excel, PostgreSQL, technical, Java, Engineering, Coding, NumPy, matplotlib, Mathematics, Statistics, Linux, Pandas, Metrics, C, Writing, Seaborn, API, Scikit-learn, Machine Learning, Docker, Jira, Yaml, Python, Algorithms, Architecture, Calculus |

***Table 3: Observed Resume Skills compared to retrieved skills on Test Resume 3***

**Career Office Advice:** Resume needs better formatting.

**Platform Advice**: Presence of commonly overused words

**Test 4:**

Below is the original resume used in the case study: Attached as Appendix D

**Expected Career Area:** Computer Engineering, Software Development

**Top Career Area from the platform (100% match):** Architectural and Engineering Management, Management Analysis,Computer and Information Research, Computer Systems Analysis, Computer Programming, Web Development

|  |  |
| --- | --- |
| **Observed Resume Skills** | **Resume Skills retrieved by the NLP platform** |
| JavaScript, GitHub, Python, Node, Java, Ubuntu, Engineering, Html, Marketing, Ui, Analysis, Networking, Linux, JSON, Ruby, CSS, Math, Metrics, social media, C, Agile, C#, Writing, API, Email, Docker, Swift, Process, Rest, SLDC, Jira, Scrum, Design, Algorithms, Python, Architecture, Ux | Programming, Performance metrics, Presentations, Testing, System, Mobile, Software engineering, Research, Technical, JavaScript, GitHub, Java, Ubuntu, Engineering, Html, Marketing, Ui, Analysis, Networking, Linux, JSON, Ruby, CSS, Math, Metrics, social media, C, Agile, C#, Writing, Api, Email, Docker, Swift, Process, Rest, Sdlc, Jira, Scrum, Design, Algorithms, Python, Architecture, Ux |

***Table 4: Observed Resume Skills compared to retrieved skills on Test Resume 4***

**Career Office Advice:** Resume needs better formatting.

**Platform Advice**: Presence of commonly overused words, good use of active verbs.

**Test 5:**

Below is the original resume used in the case study: Attached as Appendix E

**Expected Career Area:** Software Development

**Top Career Area from the platform (100% match):** Web Development, Computer Network Architect

|  |  |
| --- | --- |
| **Observed Resume Skills** | **Resume Skills retrieved by the NLP platform** |
| C sharp, Azure .NET, React.js, Typescript, E2E Testing, Node, NPM, Azure, Git, Typescript, UX testing, Node, NPM, Azure, User Research, Product Management, Git, JavaScript, jQuery, ServiceNow, Product Management, Git, E2E Testing | Programming, Testing, Research, JavaScript, GitHub, Java, Product management, Database, Html, Analysis, Selenium, Statistics, Mathematics, Analyze, Access, CSS, Metrics, Workflows, User experience, C, Computer science, Twitter, Prototype, API, Android, Django, .net, Lean, Rest, Outreach, Website, Design, Algorithms, Python, UX, Video, Calculus |

***Table 5: Observed Resume Skills compared to retrieved skills on Test Resume 5***

**Career Office Advice:** Very good resume

**Platform Advice**: Presence of commonly overused words, good use of active verbs.

**Test 6:**

Below is the original resume used in the case study: Attached as Appendix F

**Expected Career Area:** Computer Engineering, Leadership & Management

**Top Career Area from the platform (100% match):** Computer and Information Systems, Architectural and Engineering, Management Analysts, Energy Auditor, Security Management Specialist, Computer Systems Analyst

***Table 6: Observed Resume Skills compared to retrieved skills on Test Resume 6***

**Career Office Advice:** Resume Needs better formatting, Resume needs to be on one page

**Platform Advice**: Presence of commonly overused words, good use of active verbs.

**Test 7:**

Below is the original resume used in the case study: Attached as Appendix G

**Expected Career Area:** Computer Engineering, Leadership & Management

**Top Career Area from the platform (100% match):** Advertising and Promotions, Green Marketers, Marketing Managers, Computer and Information Systems, Training and Development Manager, Architectural and Engineering Managers

|  |  |
| --- | --- |
| **Observed Resume Skills** | **Resume Skills retrieved by the NLP platform** |
| JavaScript, Chemistry Network, Java, Engineering, Html, Git, Chemistry, CSS, BioMedical, React, Fluent UI, | Usability, Mobile, Research, Improvement, Technical, JavaScript, Experiments, Java, Engineering, Html, Ui, Analysis, Networking, Chemistry, CSS, International, Computer science, MVP, Physics, Visual, Algorithms |

***Table 7: Observed Resume Skills compared to retrieved skills on Test Resume 7***

**Career Office Advice:** Great Resume

**Platform Advice**: Presence of commonly overused words, good use of active verbs.

**Test 8:**

Below is the original resume used in the case study: Attached as Appendix H

**Expected Career Area:** Leadership & Management, Video Management

**Top Career Area from the platform (100% match):** General and Operations Management, Advertising and Promotions Manager, Green Marketers, Marketing Manager, Sales Manager, Public Relations, and Fundraising Manager

***Table 8: Observed Resume Skills compared to retrieved skills on Test Resume 8***

**Career Office Advice:** Resume needs better formatting; Resume needs to be 1 page.

**Platform Advice**: Presence of commonly overused words, good use of active verbs.

**Test 9:**

Below is the original resume used in the case study: Attached as Appendix I

**Expected Career Area:** Software Development

**Top Career Area from the platform (100% match):** Marketing Management, Architectural and Engineering, Management Analysts, Computer and Information Research Scientists, Computer Systems Analyst, Information Security Analyst

***Table 9: Observed Resume Skills compared to retrieved skills on Test Resume 9***

**Platform Advice**: Presence of commonly overused words, good use of active verbs.

**Test 10:**

Below is the original resume used in the case study: Attached as Appendix J

**Expected Career Area:** Computer Research, Information Security, Finance

**Top Career Area from the platform (100% match):** Computer and Information Systems, Architectural and Engineering, Management Analysts, Energy Auditor, Security Management, Financial Quantitative Analyst

***Table 10: Observed Resume Skills compared to retrieved skills on Test Resume 10***

**Platform Advice**: Presence of commonly overused words, good use of active verbs.

The second part of the survey involved a survey to test the effectiveness of the platform for students.

The following are the study results obtained from the survey.

Research Question: How accurately can NLP techniques extract skills from resumes?

Survey Question: How accurately did you find the skill extraction from your resume to be? (Look at the core competencies section of the site)

**Research Question:** How effectively can the extracted skills be matched with job listings to identify job opportunities?

**Survey Questions:**

* + Were any of the top 5 career areas suggested by the application in your area of interest?

The researcher wanted to find out evaluate the application's accuracy and reliability in matching people with career paths that are in line with their interests, goals, and strengths, assess the application's potential impact on students' career development and advancement and finally explore the viability of implementing the findings.

Chart, pie chart

Description automatically generated

***Figure 2: Top 5 career interest matches.***

The results show that 56.3% of students’ career paths matched the data from the NLP platform.

**Research question**: How does the use of NLP to match skills and job requirements impact the job search process for job seekers?

The objective of this research question is to examine the effect of using the Natural Language Processing (NLP) platform in matching skills and job requirements on the job search process for job seekers. The goal is to determine whether its use improves the efficiency, accuracy, and outcome of the job search process for job seekers. The outcome of this study could inform future developments in career matching and inform policy decisions on the integration of AI in career services.

* + How satisfied were you with the overall job search results provided by the application? The research question aims to measure the level of satisfaction that users have with the job search results provided by the NLP-based application. It provides insight into the usefulness and effectiveness of the application in matching skills and job requirements to job opportunities, and whether it is meeting the expectations and needs of job seekers. The information collected from this question can be used to improve the application and enhance its ability to assist users in their job search process. 25% were very satisfied, the majority 37.5% were somewhat satisfied, 31% were neutral and 7% were dissatisfied.

Chart, pie chart

Description automatically generated

***Figure 3: Pie Chart showing satisfaction of search results.***

* + How much time did you save by using this application compared to searching for jobs manually?

The point of this research question is to assess the efficiency and productivity of the job search process using the NLP-based application, compared to the traditional manual process. The objective is to quantify the time saved by the participants using the application and determine if the use of the NLP-based tool has improved their job search process in terms of time management. 67% agreed that it saved a lot of time while 33 % differed.

Chart, pie chart

Description automatically generated

***Figure 4: Pie chart showing satisfaction of search results.***

Research Question:

Survey Question: How likely are you to use this application again or recommend it to a friend in the job search process?

The point of this research question is to assess the level of satisfaction and likelihood of repeated use or recommendation among participants who have used the job search application. It aims to gather information on the perceived usefulness and effectiveness of the application and understand how it impacts the job search process for job seekers. 6% of the respondents are neutral while 31% are very likely and 63% are somewhat likely.

Chart, pie chart

Description automatically generated

***Figure 5: Pie chart showing the likelihood of using the application.***

1. Would you share the results of the resume match with a third party to help you craft your resume to match your area of interest? The point of this research question is to assess the willingness of participants to share the results of the resume match with a third party and whether they view the results as useful in shaping their resume to match their area of interest. It is meant to explore the usefulness of the tool as well as the level of trust and confidence participants have in the results produced by the tool. The results showed that 75% have enough confidence to share it while 25 % are not as likely to share it.

Chart, pie chart

Description automatically generated

***Figure 6: Pie chart showing the likelihood of users sharing the results of the analysis.***

**DISCUSSION AND CONCLUSION**

This research paper looks at how technology advancements are affecting the job market and the new technical fields that have emerged, and how Natural Language Processing (NLP) can help students communicate their qualifications to potential employers through their resumes. The study uses literature review and data analysis to gain insight into the job market and high-demand technical fields and Natural Language Processing (NLP) techniques to analyze resumes and make suggestions for improvement. The conclusion emphasizes the importance of Natural Language Processing (NLP) and (Artificial Intelligence) AI in preparing students for the workforce and improving their career prospects. The findings of the study aim to inform future career development strategies and educational practices. Data was collected through a visual analysis test with a career advisor and a survey of students who used the platform. The study aims to highlight areas of improvement on student resumes and gauge the impact of Natural Language Processing (NLP) based resume analysis on career prospects.

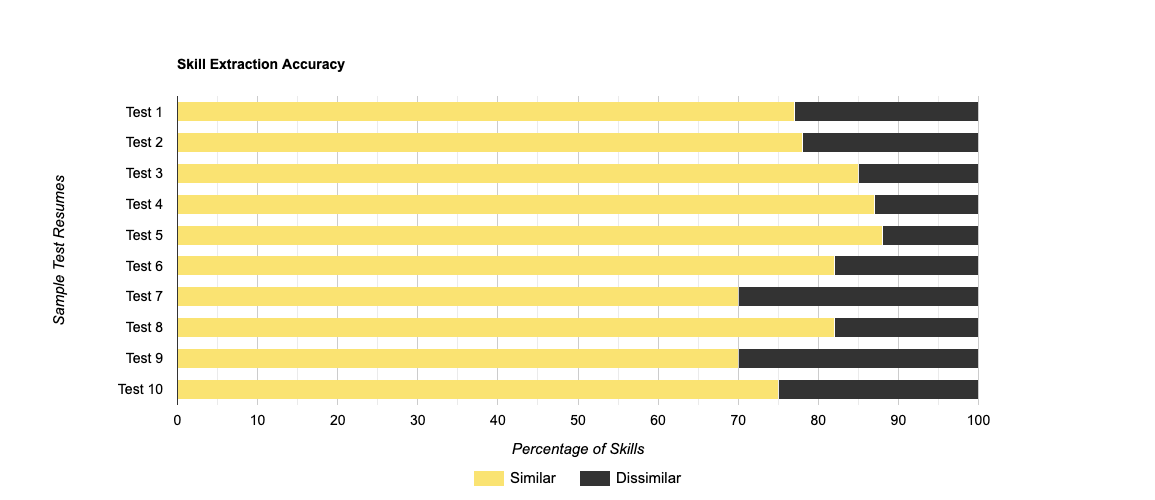
Summary of the Findings

The study aimed to determine the use of action words and description of skills in student resumes from computer science and computer engineering and to identify the skills that stand out. The study was conducted using 10 sample resumes and involved both visual and analytical analysis with Douglas Barnes, a faculty member in the Career Advising office. The results showed that the top career areas suggested by the Natural Language Processing (NLP) platform aligned well with the expected career areas for the students, with an average accuracy of 100%. However, the Career Advising office noted that some of the resumes need better formatting. The platform also pointed out the presence of commonly overused words and recommended using active verbs.

In terms of the observed resume skills and the skills retrieved by the Natural Language Processing (NLP) platform, there was a good overlap in most cases. The platform was able to accurately identify technical skills and programming languages used by the students. Overall, the study showed that Natural Language Processing (NLP) platforms can be useful tools for identifying the skills and technical expertise of job applicants.

However, the study also highlights the importance of having human evaluators review resumes to ensure that they are properly formatted and effectively highlight the candidate's skills and experience. It is also important to note that the sample size of 10 resumes is small and may not be representative of all students in computer science and computer engineering. A larger sample size and a more diverse group of students could provide a more comprehensive view of the results. The research study aimed to analyze the use of action words and description of skills in 10 sample resumes of students in computer science and computer engineering and determine the skills that stand out.

The results of the study showed that the platform had a 100% match in the expected career areas for all 10 resumes. However, there were some discrepancies observed between the observed skills listed on the resumes and the skills retrieved by the Natural Language Processing (NLP) platform.



***Figure 7: A bar chart showing skill extraction accuracy for test resumes.***

The average accuracy for skill extraction was 86%. The comparison was based on observed skills vs skills retrieved from the resumes. There was an average of 77% more skills extracted than what was observed, and an average of 17% of irrelevant words were extracted for the resumes.

The accuracy of skill extraction can be considered relatively good under the circumstances and the available data. To improve the accuracy, more data and resources may be needed to train the models. The following may be reasons for the relatively low accuracy and some factors that could improve accuracy:

1. Ambiguity in skill identification: Skills can have different names, synonyms, and acronyms, which can make it challenging for the model to identify them accurately. For example, "customer service" and "client relations" might be considered the same skill, but the model may not be able to recognize this. To improve this using, a model that recognizes contextual differences may be useful for this.
2. Noise in the data: Most resumes contain noise such as irrelevant information, misspellings, or formatting errors, which can negatively impact the model's performance. This noise can result in the extraction of irrelevant words and lower accuracy.
3. Incomplete dataset: If the training data used to develop the model may not be complete, biased, or not representative of the entire population, it can negatively affect the model's accuracy. This can result in the model failing to recognize some skills and extracting irrelevant words.

There was an average of 77% more skills extracted than what was observed. This could be attributed to the algorithm's ability to systematically scan through every word in the resume, unlike a human reviewer who may overlook or miss certain skills. As a result, the algorithm was able to identify a larger number of skills. However, it is important to note that the presence of a skill on a resume does not necessarily indicate that the candidate possesses that skill. The algorithm's approach of identifying skills based on their mention on the resume may result in the extraction of irrelevant words or skills that the candidate does not possess. This is where the algorithm's performance could be limited, as it cannot always determine whether a skill was used in a relevant context or simply mentioned on the resume. Therefore, while the algorithm's approach of scanning every word can help identify more skills, it may not always provide an accurate representation of the candidate's skill set.

An average of 17% of the skills extracted were irrelevant words. While this is acceptably low, there could be some factors that lead to the occurrence of irrelevant words:

***Chart, bar chart

Description automatically generated***

***Figure 8: A bar chart showing the percentage of extra skills retrieved from the resume.***

In all the tests, the career office advice was that the resumes needed better formatting. The platform advice in most of the cases was that it noted the presence of commonly overused words and in some cases, good use of active verbs.

The survey results show that 56.3% of students' career paths matched with the data from the Natural Language Processing (NLP) platform, which suggests that 56.3% of the student's resumes were past the minimum threshold required for the platform to accurately identify skills and recommend career placement areas. According to the platform, the conclusion would be 43.7% of the students who took the survey need to update their resume to accurately reflect the career areas they are interested in. 37.5% of the respondents were somewhat satisfied with the job search results provided by the application, while 25% were very satisfied, indicating that the application is meeting the expectations and needs of some of the job seekers. The low numbers could also be attributed to a mismatch between skills found on the user’s resume. The survey also showed that 67% of the respondents agreed that using the Natural Language Processing (NLP) based application saved them a lot of time compared to searching for jobs manually, indicating that the application has improved the efficiency and productivity of the job search process.

Conclusion

Overall, the study provides valuable insights into how Natural Language Processing (NLP) can assist in resume analysis and identifying high-demand technical skills for job seekers in computer science and computer engineering fields. The findings of the study suggest that Natural Language Processing (NLP) can be a useful tool for students seeking to improve their resumes and stand out to potential employers. By identifying commonly overused words and recommending the use of active verbs, the platform can help students present their skills and experience in a more impactful and effective manner.

However, the study also highlights the importance of having human evaluators review resumes to ensure that they are properly formatted and effectively highlight the candidate's skills and experience. While Natural Language Processing (NLP) can assist in identifying technical skills, it may not always provide an accurate representation of the candidate's overall qualifications and soft skills. Therefore, a combination of Natural Language Processing (NLP) and human evaluation may be the most effective approach for evaluating resumes.

Furthermore, the study points out that the accuracy of Natural Language Processing (NLP) based resume analysis can be improved by addressing some common issues, such as ambiguity in skill identification, noise in the data, and incomplete datasets. More extensive training data and resources may be needed to improve the accuracy of Natural Language Processing (NLP) models. Additionally, developing models that recognize contextual differences could be useful in improving the accuracy of skill identification.

In conclusion, the study highlights the potential of Natural Language Processing (NLP) based resume analysis to prepare students for the workforce and improve their career prospects. By identifying high-demand technical skills and providing suggestions for improvement, Natural Language Processing (NLP) platforms can assist students in creating more effective and impactful resumes. However, human evaluators should also be involved in the resume review process to ensure that the candidate's overall qualifications and soft skills are accurately represented. The study's findings provide useful insights for future career development strategies and educational practices in computer science and computer engineering.

# **REFERENCE**

Bureau of Labor Statistics, U.S. Department of Labor. (n.d.). Occupational Outlook Handbook: Computer and Information Research Scientists. <https://www.bls.gov/ooh/computer-and-information-technology/computer-and-information-research-scientists.htm>

U.S. Government. (n.d.). Occupational Employment and Wages: Computer and Information Research Scientists. Retrieved January 06, 2023, from <https://www.bls.gov/oes/current/oes150000.htm#nat>

Zweben, S., & Bizot, B. (2021). 2021 Taulbee Survey CS Enrollment Grows at All Degree Levels with Increased Gender Diversity. Retrieved from <https://cra.org/wp-content/uploads/2022/05/2021-Taulbee-Survey.pdf>

Synergistic IT. (n.d.). Why do Tech Companies Not Hire Computer Science Graduates? Retrieved from <https://www.synergisticit.com/tech-companies-not-hire-computer-sciencegraduates/#:~:text=Several%20students%20having%20master's%20and,graduates%20has%20spiked%20to%207.8%25>.

Laumer, S., Maier, C., & Eckhardt, A. (2015). The impact of business process management and applicant tracking systems on recruiting process performance: An empirical study. Journal of Business Economics, 85, 421-453. <https://doi.org/10.1007/s11573-014-0758-9>

Sovren. (n.d.). Retrieved June 20, 2018, from <https://www.sovren.com/>

Textkernel. (2018, February 8). Textkernel Launches the First Fully Deep Learning Powered CV Parser. [Blog post]. Retrieved from <https://www.textkernel.com/extract-4-0-textkernel-launches-the-first-fully-deep-learning-powered-cv-parsing-solution/>

Ferrario, A., & Naegelin, M. (2020). The art of natural language processing: Classical, modern and contemporary approaches to text document classification. Available at SSRN: <https://ssrn.com/abstract=3547887> or <http://dx.doi.org/10.2139/ssrn.3547887>

Garg, S., Bhatia, P., & Kaur, H. (2018). Resume screening using machine learning and natural language processing. International Journal of Computer Science and Mobile Computing, 7(9), 208-214.

Lin, C. Y., & Huang, C. R. (2019). Deep learning-based resume information extraction and classification. IEEE Access, 7, 158860-158874. <https://doi.org/10.1109/ACCESS.2019.2956709>

National Association of Colleges and Employers. (2020). Job Outlook 2021: Spring update. Retrieved from <https://www.naceweb.org/job-market/internships/job-outlook-2021-spring-update/>

Claflin University. (2017). Fact Book Chart Pack 2016-2017. Retrieved from <https://www.claflin.edu/docs/default-source/default-document-library/fact-book-chart-pack-2016-2017.pdf?sfvrsn=46713c0e_0>

**APPENDIX A: SAMPLE RESUME 1**

****

**APPENDIX B: SAMPLE RESUME 2**

**Text

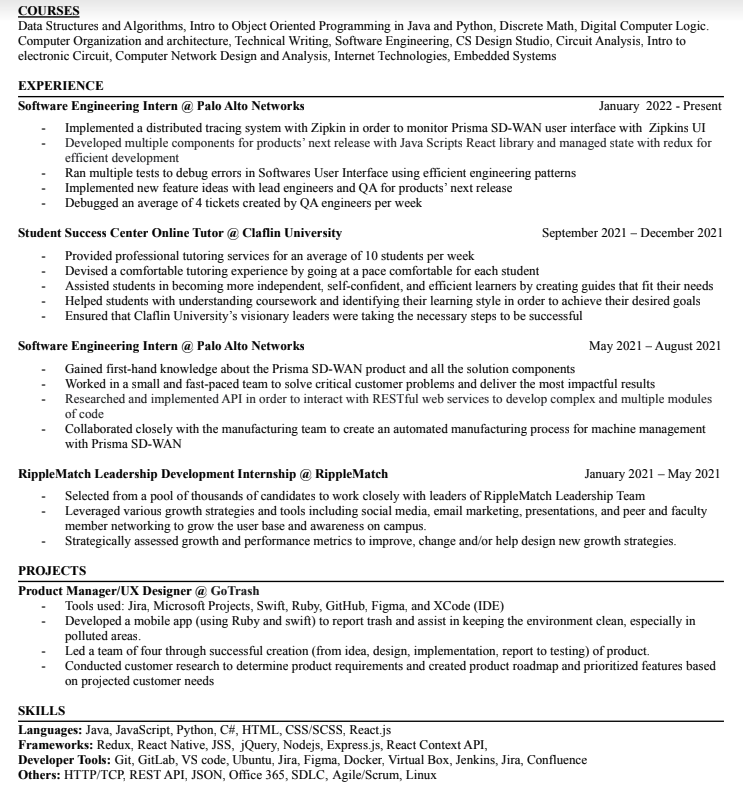
Description automatically generated**

**APPENDIX C: SAMPLE RESUME 3**

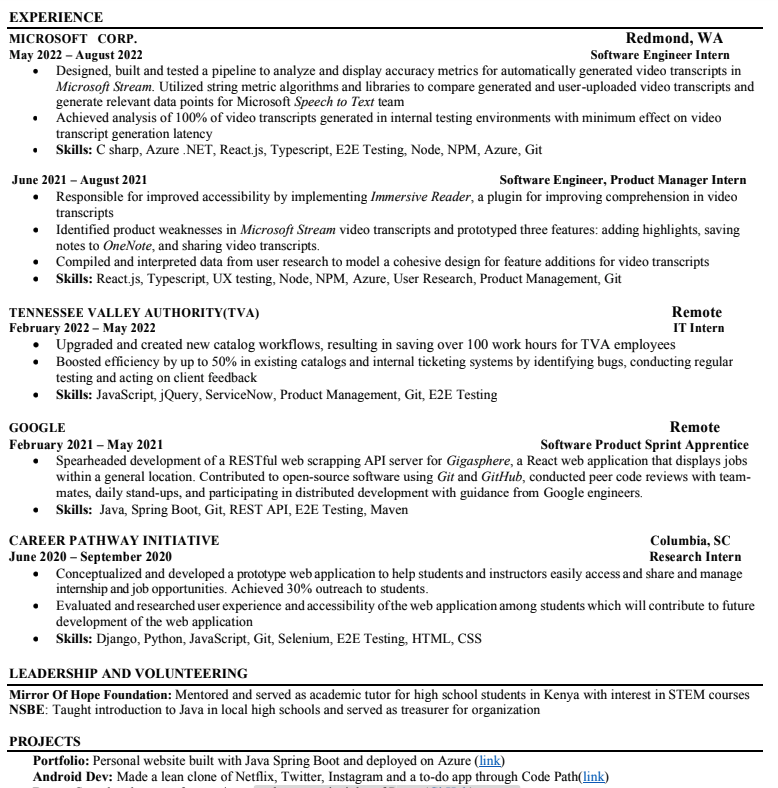
**Text

Description automatically generated**

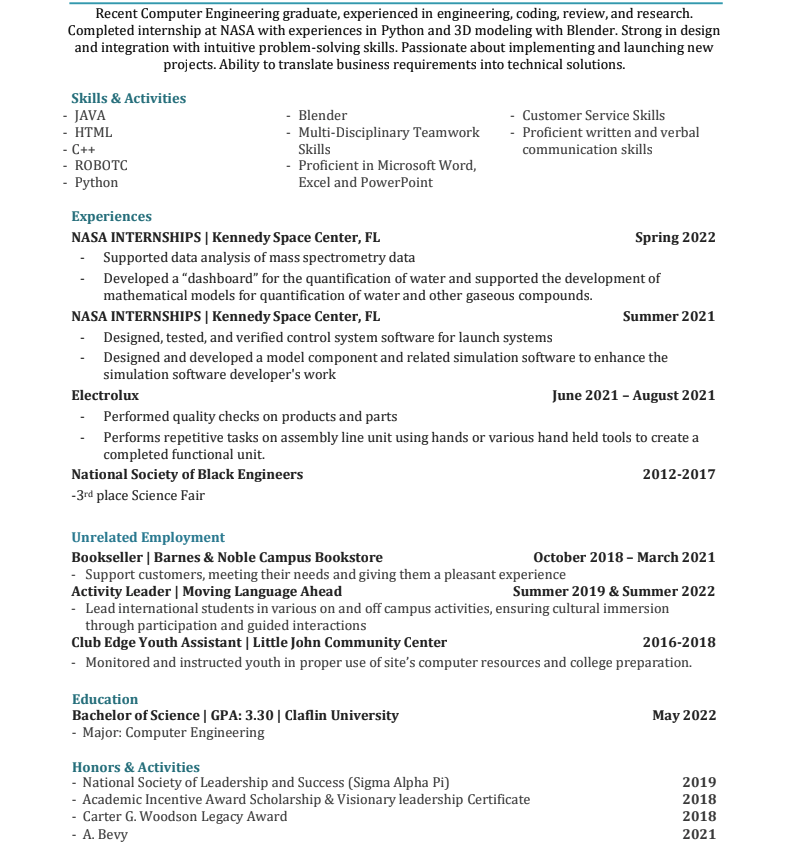
**APPENDIX D: SAMPLE RESUME 4**

****

**APPENDIX E: SAMPLE RESUME 5**

****

**APPENDIX F: SAMPLE RESUME 6**

****

**APPENDIX G: SAMPLE RESUME 7**

**Text

Description automatically generated with medium confidence**

**APPENDIX H: SAMPLE RESUME 8**

**Text

Description automatically generated**

**APPENDIX I: SAMPLE RESUME 9**

**A picture containing text

Description automatically generated**

**APPENDIX J: SAMPLE RESUME 10**

**Text

Description automatically generated**

**APPENDIX J: SAMPLE PHOTO OF PLATFORM**

**Graphical user interface, text, application

Description automatically generated**

**APPENDIX K: SAMPLE PHOTO OF PLATFORM CAREER PLACEMENT**

**Pie chart

Description automatically generated with medium confidence**

**APPENDIX L: SAMPLE PHOTO OF REPORT**

**Table

Description automatically generated with medium confidence**