**Date of form completion: Saturday, February 5 2022**

**Student name: Joshua Kiprono**

**Student’s possible graduation date: May 2023**

**Students required courses remaining for graduation:**

**Research category: Project**

**Main adviser's name: Shrikant Pawar**

**Committee faculty name 1:**

**Committee faculty name 2:**

**Committee faculty name 3:**

**Note: Answer all the questions in as much detail as possible. If you need help, please contact your adviser or the committee faculty member for help.**

--------------------------------------------------------------------------------------------

**Section 1: Introduction and literature review.**

**Diagram

Description automatically generated**

1. **What is your established area to research?**

Using machine learning to improve outreach of career opportunities.

1. **What is the research question you want to address?**

**Does a web application using Natural Language Processing increase the rate of students finding opportunities in their field of study?**

1. **Is research within the achievable research goals? Please explain how?**

The research is within achievable research goals. The goal of the research is to investigate how an application using Natural Language Processing can be a tool to improve how students find opportunities through their instructors and fellow students. Machine learning is the study of algorithms that can improve over time while extracting information from data. The goal is to use these algorithms to help Claflin gauge how students interact with opportunities while also increasing accuracy of opportunities that students receive based on their majors.

1. **Have you considered argumentative, integrative, historical, methodological, systematic, and theoretical sections?**

**Argumentative section**: The Office of career development is not achieving optimal outreach to students concerning career opportunities nor are they finding clear metrics about student interactions about the opportunities sent out to students.

**Theoretical section**: Part of the areas machine learning will be used is text analysis using Natural Language Processing. I will be using context analysis to break down sentences in emails to extract n-grams, noun phrases, themes and other elements of text. I will then run these against a pre-trained model to extract emails that to do with career opportunities and structure it in a way that can be displayed to the student in a clear and easy to follow manner.

1. **What is your problem statement?**

Claflin’s Office of Career Development sends hundreds of mails to students concerning career opportunities. However, there is no standardized metric for the school administration to find how many students follow up with applying to these opportunities. The goal of Office of Career Development is to help as many students as possible achieve their career objectives yet there is no established system to help the office make decisions based on real time data. On the other hand, tracking and managing several applications can be a tedious process for students as the student has to find an easy and efficient way to manage applications. Not all productivity software and calendar applications are built in a way that enhances productivity and efficiency. In this research, I will be investigating how to use machine learning to increase the effectiveness of the Office of Career Development in sending career opportunities and for students to increase their productivity while finding career opportunities.

1. **State existing literature on selected topic?**

Harrington, P. (2012). Machine learning in action. Manning.

Bird, S. (2016). Natural language processing with python. O'Reilly Media.

1. **What is existing data evaluation, analysis and interpretations available?**

In an initial test website designed to tackle the problem statement:

“Tracking and managing several applications can be a tedious process for students as the student has to find an easy and efficient way to manage applications. Not all productivity software and calendar applications are built in a way that enhances productivity and efficiency. In this research project, I will be looking into creating a user-focused application manager that will promote efficiency in managing and tracking applications.”,

An online survey was among students who had used the site. The survey form was sent to 16 students of whom responded.

The responders’ majors are Biology, Math, Computer Science and Computer Engineering (12.5%, 12.5%, 50% and 12.5% respectively). 50% of the responders used laptops to access the site

100% of users responded that they would use an online service to track and manage their applications.

Some of the recommendations given includes adding a section for prospective graduate students, improving user interface and sourcing opportunities for users.

This survey helped create a foundation in this research experiment that extends the capabilities of the initial web application to include instructors to share opportunities to a wide group of students as well as the office of career development.

1. **Are the articles credible, and address objectivity, persuasiveness and values of conducted study?**

The articles mentioned are credible and address objectivity, persuasiveness and values of the conducted study. The articles discuss the core concepts of machine learning, training models, NLP and problem solving with app development and machine learning

**Section 2: Methods.**

1. **Introduce the overall methodological approach for investigating research problems?**

This research follows an experimental design. The following general methodological approach will be used to investigate the problem:

1. Consider variables and how they are related. In this case, the dependent variable is using an application to connect the Office of Career Development and students. The independent variable is students efficiently finding related career opportunities.
2. Writing a specific, testable hypothesis. Does using the application increase the efficiency of instructors, Office of Career Development and students in finding and disseminating opportunities to students?
3. Design and create a test application to see any effect on the independent variable.
4. Assign subjects to groups. The groups will be students using the application and those not using the application.
5. Measure the results and come to conclusion.
6. **Indicate how the approach fits the overall research design?**

The research follows an experimental design. The design involves creating an application (Creating a procedure) to test the hypothesis that a Natural Language Processing algorithm can be used to increase the efficiency of students finding career opportunities and the Office of Career Development in finding an accurate metric to gauge number of students looking for opportunities.

1. **Describe the specific methods of data collection you are going to use?**

Conducting online surveys – Using a third-party service to create surveys and sending them to students

Online Tracking – Once the app is created, I can use analytics to view how users interact with the site and track productivity and usage statistics which can be utilized in coming to research conclusions.

Interviews – Asking instructors and The Office of Career Development statistics on career field alumni have ended up in.

1. **Explain how you intend to analyze results?**

Quantitative approach: Quantify data and generalize the collected data.

Qualitative approach: I intend to put the data into context and interpret the data gathered from students and the career office to investigate research problems.

1. **Provide background and rationale for methodologies that are unfamiliar for the readers?**

An experimental methodology was used as I am studying causal relationships. This means that a change in one variable causes a change in another variable. The variables in this case are a test web application and a change in efficiency of staff and students in finding and sharing career opportunities.

1. **Provide a rationale for subject selection and sampling procedure?**

The study size will consist of approximately 20 students and staff. The study subjects will then be subjected to random assignments to three groups: The control group, a group that uses the application and a group that does not use the application (i.e uses the conventional method of finding career opportunities).When assigning subjects to groups, I intend to use a randomized block design where subjects are randomly grouped into staff or students.

1. **Address potential limitations?**

The sample size depends on the number of research subjects willing to do participate in the test.

The test is dependent on an application that is expected to work perfectly as theorized.

Section 3: Results.

1. State an introductory context for understanding the results by restating the research problem that underpins the purpose of study?
2. Summarize the key findings arranged in a logical sequence?
3. Include non-textual elements, such as figures, charts, photos, maps, tables, etc. to further illustrate the findings?
4. Summarize the results, highlighting for the reader observations that are most relevant to the topic under investigation?

Section 4: Discussion.

1. State the major findings of your study?
2. Explain the meaning of the findings and why are they Important?
3. Relate the findings to similar studies?
4. What are alternative explanations of your findings?
5. What are your study limitations?
6. What are suggestions for further research?

Section 5: Conclusions.

1. Present the last word on the issues raised by you?
2. Summarize thoughts and convey the larger implications of your study?