

**DEDAN KIMATHI UNIVERSITY OF TECHNOLOGY**

**Project Title:**

**MACHINE LEARNNG MODEL TO SHORTLIST CANDIDATES FROM JOB APPLICATIONS.**

**FAITH CATHERINE OTIENO**

**C026-01-0306/2018**

A project submitted to the Department of Computer Science in the School of Computer science and IT in partial fulfillment of the internal attachment in Computer Science.

January 2021.

DECLARATION

This proposal is my original work and has not been presented for a degree in any other University

Name: ………………………………………………………………………………………..

……………………….…. ………………………………

Signature Date

This proposal has been submitted for examination with my approval as University Supervisor

Name: ………………………………………………………………………………………..

……………………… ……………………………………

Signature Date

**ABSTRACT**

The number of job applications has increased over time and is statistically said to increase from 10% to 20% in certain sectors such as business and medicine. Throughout times reviewing of applications has been done manually by human beings. This work continues to be tedious as the number of applications to be reviewed continues to increase. This system aims to help solve this problem by scanning the resumes and short listing the best fit candidates using a machine learning model. The system will be built using a waterfall approach. The system will be trained using a resume dataset and also tested to ensure its function with limited errors. This system is highly recommended as it will ease the burden of scanning through applications and also ensure companies get their best fit candidates according to their requirements.

Table of Contents

[Chapter 1 INTRODUCTION 1-6](#_Toc62938815)

[BACKGROUND 1-6](#_Toc62938816)

[PROBLEM STATEMENT 1-7](#_Toc62938817)

[Objectives 1-7](#_Toc62938818)

[Main objective 1-7](#_Toc62938819)

[System objectives 1-7](#_Toc62938820)

[Justification. 1-8](#_Toc62938821)

[Scope 1-8](#_Toc62938822)

[Target Users. 1-8](#_Toc62938823)

[Platform and where to implement 1-8](#_Toc62938824)

[Chapter 2 : LITERATURE REVIEW 2-9](#_Toc62938825)

[INTRODUCTION 2-9](#_Toc62938826)

[Talent Sourcing. 2-9](#_Toc62938827)

[Talent Recruitment. 2-9](#_Toc62938828)

[Candidate Screening and Engagement. 2-9](#_Toc62938829)

[CASE STUDIES 2-10](#_Toc62938830)

[Case study 1: IDEAL’s virtual assistant 2-10](#_Toc62938831)

[Case study 2: Avrio AI Inc. 2-11](#_Toc62938832)

[Case study 3: Entelo 2-11](#_Toc62938833)

[Case study 4: Mya Systems. 2-12](#_Toc62938834)

[Summary 2-13](#_Toc62938835)

[Research Gap 2-13](#_Toc62938836)

[Proposed methodology 2-13](#_Toc62938837)

[Chapter 3 :METHODOLOGY 3-14](#_Toc62938838)

[Introduction 3-14](#_Toc62938839)

[Fact finding techniques 3-14](#_Toc62938840)

[Record Views 3-14](#_Toc62938841)

[Observation 3-14](#_Toc62938842)

[Software design - Software development procedures. 3-14](#_Toc62938843)

[Requirements 3-15](#_Toc62938844)

[User Requirements 3-15](#_Toc62938845)

[System Requirements 3-15](#_Toc62938846)

[Non-Functional Requirements 3-15](#_Toc62938847)

[Preliminary Data Processing and analysis. 3-15](#_Toc62938848)

[REFERENCES 3-15](#_Toc62938849)

# INTRODUCTION

## BACKGROUND

Growth of the population has seen an increase of job seekers especially among the youth. These numbers are estimated to rise in the coming years as society sees millions of entries to universities, colleges and tertiary schools globally, this excluding the number of people that will seek manual labors and other sorts of work. . Certain areas such as technology, medicine and business show more increase in job seekers with areas in technology estimated to increase up to 30%. These numbers will increase even as society increases job opportunities to curb unemployment in society.

The process of hiring in companies is a series of different activities. Companies will first have job postings specifying the job qualifications, deadlines of applications and documents to be attached. Once the posting is done and applications are received, the human resource department of the company will go through the resumes to find the applicants with the needed qualifications. Depending on a set target of applicants needed to be interviewed, the company then goes ahead to shortlist the candidates seen to be best fit. These are the candidates that will be called for an interview. This process maybe slightly different depending on the company’s policies.

In the era of technology, majority of job postings and job applications are done online. Some employers find their employees in online platforms such as linked in. Many employees build portfolios and profiles online that can be evaluated by the employers. Unlike olden days, most applications are sent by email or through companies platforms and employers communicate with the applicants through this same platforms, some interviews have even been successfully conducted online.

Many companies have seen increase in number of job applications as job seekers increase. On average, corporate job posts will receive up to 250 applications. Most companies will want to select 4 to 6 of the candidates for an interview. Therefore most companies will need to read through resumes and eliminate more than 200 resumes so as to remain with candidates they consider a fit with their company’s requirement.

Companies face many challenges in the process of short listing candidates best fit for the job description. These challenges involve reading through the many applications sent, matching skills with the ones required in the job description, doing background checks and verification on external activities done by the applicants such as checking their profiles, checking projects done and checking volunteer work.

As the human abilities are limited, it takes far too much time to shortlist candidates. The process of reading all those applications is challenging and quite exhausting for the department. According to a research done it takes an average of 29 to 35 days for companies to short list candidates and call the desired candidates back for an interview. This therefore wastes a lot of time as a more than a month is required before the company can hire a candidate for a job application.

Many companies have also complained of the bulk of resumes with no skills or very little skills. Many companies face challenges of shortage of expertise even from the applications they receive therefore, a lot of work is used in scanning resumes that are not valuable for the job description, therefore making the process more tiring and discouraging as it is hard to find the candidates best fit for the job. The process of short listing candidates is therefore very time consuming and exhausting when done by humans.

This has therefore brought a need for better, easier and more efficient ways to shortlist candidates from a pool of applications. This has therefore brought about building of models that will be part of the process of hiring either through suggesting candidates for a company to look into from profiles, reading resumes and matching skills with those in the job description or models that are part of the interviewing process.

## PROBLEM STATEMENT

Companies have a difficult time in reading through job applications to extract skills and match them to the skills in the job description due to very many job applications. Companies also find it difficult to filter through the applications and choose the best fit for an interview.

## Objectives

### Main objective

To build a machine learning system that evaluates applications and shortlist candidates best fit for the job.

### System objectives

1. To evaluate candidate resumes.
2. To find out the candidates with minimum qualifications for the job.
3. To find out candidates with the preferred job qualifications.
4. To short list the best candidates viable for an interview.

Research Questions

1. Are there uploaded resumes to evaluate?
2. Are there applicants with minimum job qualifications?
3. Are there applicants with preferred job qualifications?
4. Are there candidates considered to be best fit for the job?

## Justification.

The main reason for creating this system is to provide companies with an easier way to filter through applications and shortlist best fit candidates therefore making the hiring process faster and easier for the companies.

By having intelligent models review resumes and make decisions of the best fit candidates, the workload is reduced and resources that were to be used for this process are diverted to other areas. The company also gets to save on time in the hiring process as the machines will use a shorter time to review and make decisions as compared to humans.

This system is also a necessary venture to avoid selecting unqualified candidates, that is candidates with zero to no skills.

## Scope

## Target Users.

This system will have two types of users. The first user is the company. The company will post a job description giving specific skills and requirements they need. The second user is the applicant who on viewing the job posting will be able to apply by uploading their resume and any other documents they have been told to attach. The system will then process the resumes and give the company a feedback of shortlisted candidates.

## Platform and where to implement

The system is a machine learning model that will be embedded in a web service. The web interface will show all the job postings and allow for uploads. The machine learning model will be built using python and the web interface using the web languages, HyperTextMarkupLanguage(HTML), Cascading Style Sheets(CSS) and JavaScript.

# : LITERATURE REVIEW

## INTRODUCTION

This review aims to show already existing application tracking systems and the approaches they have taken to help in the process of hiring. Artificial intelligence can be helpful to the hiring process in different ways, models in the market have been useful in talent sourcing, talent recruitment and candidate screening and engagement.

### Talent Sourcing.

Talent sourcing is the process of looking for talent needed by companies. AI models have been built to assist in talent sourcing. This process may involve predictions on candidates who may be looking for a new job in the near future that is depending on their posts or the state of companies they currently work for. Models will look for such candidates matching them with the company. Models also through looking at platforms such as Linked in and Zety, they match job seekers that are more likely to qualify for positions with the company.

### Talent Recruitment.

Talent recruitment involves assisting in the repetitive processes of the hiring process such as resume and application review. Such models go through applications sent by candidates to extract the skills and qualifications and match them to the job description of the company. The candidates more qualified for the job will be recommended to the company while those with little or no skill will be left out. This eases the burden of reading through all those applications.

### Candidate Screening and Engagement.

Candidate screening and engagement involves having conversations with the candidates. Most of these models use natural language processing and can be helpful in inquiries by answering candidate questions or being part of the interview process. Some models are able to break down answers given by candidates and analyze them to make conclusions on social skills like communication and team work skills and also on virtues needed such as integrity and commitment. Some models can even have video calls with candidates and have abilities to analyze the candidates from their posture, intonation and other gestures.

All these are ways in which Artificial Intelligence has been used to assist humans in the process of hiring. All these models are very helpful in reduce the work load, saving time and choosing the best fit candidates.

## CASE STUDIES

### Case study 1: IDEAL’s virtual assistant

The ideal’s assistant integrates with the client’s existent applicant tracking systems and is trained on the client’s data set. This ensures the model is specific to the client’s needs as it is aware of their needs and their previous hiring processes. It is a model that has already been trained on more than a million hiring decisions and is therefore very efficient in decision making as it knows how to filter and make the best choices given a certain type of circumstances. This also allows the model to quickly adapt to the new clients process of hiring.

The ideal platform delivers two types of services.

**Resume screening**

The platform provides resume screening helping clients. Resume screening process done by different is quite different from what is done manually and creates inherent bias. The model by stating parameters that are standard to the company, state what an ideal candidate would look like for the company and then the model uses pattern recognition to identify those elements in applications. The model can then eliminate applications that don’t fit the criteria and remain with the candidates best fit for the job.

**Candidate Sourcing**

Ideal has claims that its virtual assistant can do candidate sourcing for companies. By training using the companies datasets and learning their preferences, the model has the ability to connect to external third party websites such as career builder and then searches these platforms to match candidates matching the employers criteria.

The ideals virtual assistant is a very useful model in the hiring scene. The algorithm performs two services that are very essential in the process of selecting candidates. It assists in both talent sourcing and talent sourcing. The team from ideal has reports that the model being used in Canada’s largest bookstore chain reduced their cost per hire by over 71% and tripled their hiring of qualified candidates.

The gap present in the model is very minimal as the only service that is not present in this model is that of candidate engagement. The model does not provide any assistance in the process of inquiries or interviews, thus no assistant in choosing the final candidates that will be employed.

The similarity of this model and the short listing system is that they both assist in talent recruitment through resume screening. This model is different from the short listing system as the short listing system will not provide any services on talent sourcing.

### Case study 2: Avrio AI Inc.

The Avrio’s main distinctive feature is in the realm of candidate engagement. Apart from features in candidates sourcing and recruitment, this model claims to incorporate a chatbot from Facebook to chat with the candidates in a pre-screen interview format.

The assistant Rio asks the candidates different questions to gauge on their knowledge concerning the position they applied for. Some of the questions may include how they will handle certain scenarios and challenges that may occur in the work place. The model is able to analyze the candidate’s answers to see their knowledge skills and abilities to work in the job therefore making decisions on the more efficient candidates. Rio can also quote salary ranges based on the candidates social media and the company’s salary range and therefore identifying candidates more qualified in matters of salary.

This model uses natural language processing to analyze tits conversations with candidates and then uses the analyzed data to inference decisions. This model is very efficient in engaging candidates pre-interviews to find candidates most qualified for the job while eliminating those with lower skill levels and those that don’t meet the company’s requirement or job description.

This model is very efficient as apart from talent sourcing and talent recruitment, it goes a step farther into doing pre interviews so as to assess candidates. The gap present in this model is that it is not involved in the actual real interview process. It is otherwise a very efficient model as it filters candidates in three processes thus allowing for the best fit candidates.

This model is very different from the short listing model. While the short listing model has only the resume screening process, The Rio model goes way further into looking for candidates viable for the job and further into doing pre-interviewing candidates that it has found to be more qualified for the job, so as to narrow down the candidates using salary ranges and other job skills that the company may require.

### Case study 3: Entelo

Entelo is a platform with a machine learning model for talent sourcing. The company uses Artificial Intelligence to help companies find candidates viable for a job. The company claims that its algorithm is capable of identifying candidates who have a 30% likelihood of changing their jobs within the next 90 days.

The Entelo model has identified more than 70 variables that they use for prediction. These variables are used to analyze data from candidates or their current employers and then make decisions based on these variables. The model will then suggest to the company employees who are more likely to be looking for a new job, therefore providing the company with talent they needed.

Some of the variables that the algorithm looks into include, update in candidates profiles such as their LinkedIn profiles. For example, the candidate may include posts such as “coming to the end of my contract” or “ time to look for new projects”. Such updates on a person’s social profile may be a sign that the person will be leaving their current place of work and may be looking for a new job in a different place.

The algorithm may also look into the health of the company where a candidate works. Any indications of layoffs or stock fluctuations in a company may be signs of the candidates losing their jobs and therefore higher chances of them looking for a new job. Layoffs may cause the candidate to lose their job and stock fluctuations may lead to the company the candidate is currently working in to close. Both scenarios will lead to the candidate looking for a new job.

The Entelo platform looks for passive candidates, that is candidates that are still in other employments and are not currently looking for work but are still open to new opportunities. The platform will therefore find these candidates and recommend them to companies that may be looking for talent with their qualifications. The platform claims that it provides the companies with access to more than 275 million passive candidates.

In a case study the platform report of their client Opower an oracle company was able to improve its percentages of hiring more qualified candidates. The company increased its hire of female candidates from 40% to 47% and their minority technical hires from 1.5% to 11%. This study therefore proving increase in efficiency of hires according to the clients criteria by using this platform.

This model is quite different from the short listing model as it doesn’t do any resume screening but only does technical outsourcing. This model is still an efficient system in the hiring process.

Case study 4: Mya Systems.

Launched by the Mya Systems in July 2016, The Mya Artificial Intelligence Recruiting system uses natural Language processing and deep learning frameworks to analyze dialogues with candidates. The company claims that its model incorporates two main methods in its candidate engagement process.

**Entity Extraction**

This involves extracting and organizing information from text into categories like locations, experience, education and skills.

**Sentence semantic analysis**

This is a method of text interpretation by looking for the similarities in sentences and thus finding interpretation. This method is very useful in optimizing searches and categorizing elements.

Mya integrates into the clients applicant tracking system and immediately starts to communicate with candidates who have applied for a job. The model then asks follow up questions to gauge knowledge on the depth of skills of the applicant. The model also scores applications based on data extracted from their resumes and other attached documents.

The platform has claimed that from applications ranging up to 4000 the system is able to reduce the time required to hire a candidate by more than 70% and that more than 91% of their candidates complete the screening. This platform is therefore quite efficient in helping in the process of hiring.

The major gap in this model will be in the talent sourcing but otherwise itrs processes are very efficient in the hiring process. This model is quite different from the short listing model as it is mainly based on candidate engagement to make decisions while the short listing system is involved in talent recruitment through reviewing applications.

### Summary

Most of the artificial Intelligence system used in the hiring process assists using one or two of the main methods used in the hiring process. These models are quite efficient as they train using the client’s datasets therefore solving issues per client and also as they have been trained in making decisions of a wide range, most of the decisions made are highly accurate. Most of these models have reports of successes with the companies they have worked with showing increase percentages in hiring more qualified clients, reduction of resources used in the hiring process and less time used in the hiring process. The models therefore make it very efficient for the companies that use them.

### Research Gap

The major research in the talent recruitment process that is application reviewing is the steps taken before the short listing process is completed. Most of the models that use this method only takes one step, they check the skills that a candidate has and match them with the company. This approach may not necessarily find the best fit candidate. The models should narrow down the decision making into three stages thus allowing for the best fit candidates. The three stages should include, classifying candidates with minimum qualifications and those without any qualifications. This will eliminate resumes with no qualifications. The next step is to then classify those with preferred qualifications and the last will be to shortlist the best fit candidates for the job.

### Proposed methodology

The short listing machine learning model aims to use the K Nearest Neighbor classification algorithm to find resumes with skills very close to the required candidate according to the company’s description. The process will involve three steps to filter the resumes, that is minimum qualification, preferred qualification and then short listing**.**

# :METHODOLOGY

## Introduction

This chapter comprises of the data collection methods aimed in carrying out the process of project development and also the software development life cycle, they include the software models and other documentation. Designing of the systems aimed to focus on the users’ need and ensuring the system is user friendly and interactive. The waterfall model was concluded to be most suitable during the software development phase. Progress flows in a downward manner for this development cycle model.

## Fact finding techniques

In conducting my research, there was need to collect facts and all relevant information. The facts when expressed in qualitative form are termed as data. The success of any project is depended upon the accuracy of available data. Accurate information can be collected with help of certain methods/ techniques. These specific methods for finding information of the system are termed as fact finding techniques. Record Views and Observations are the different fact finding techniques used by the analyst. In this study, the outlined are some of the fact finding techniques implored.

### Record Views

There was necessity to go through information related to the system that is published in sources like newspapers, magazines, journals, documents etc. The record review helped me to get valuable information about the already existing systems and the users and why there is a need for this system. Looking into current companies and the issues they face in hiring candidates. Online research was also done on the related productions.

### Observation

Unlike the other fact finding techniques, in this method the analyst looks into how companies manually do the hiring process and observes and understand the flow of documents, working of the existing systems if there are any and the users of the system.

## Software design - Software development procedures.

The approach set to be used in the system development methodology is the Waterfall Design. This is the implementation of SDLC stepwise into the system development methodology. The waterfall model was considered because it is simple to use and also the needs and requirements of the user had been clearly understood.

***The phases involved are as follows:***

**Planning**

Here in developing of the new system the first step is to identify a need for the *system*, and also plan how to develop the functional requirements of a system. This includes conducting a feasibility study to determine developing a project plan and estimating the viability of the system in question.

**Analysis and specification**

Here I seek to analyze the current systems and investigate any problems associated with it. I also seek to look at other sources of information about this system and the new requirements that concerns this system. In this phase I’m aiming to establish whether the system is viable to implement and possible problems it would solve. Moreover, I’ll need to establish that the requirements for the system development are available timely

## Requirements

### User Requirements

* A system for companies to post job posts.
* A system that allows companies to specify the job requirements.
* A system for applicants to upload resumes for a job post.
* A system to shortlist the most viable candidates for a job opening.

### System Requirements

* Functional Requirements
* Enable companies to create an account.
* Allow organizations to log in using a password.
* Enable organizations to post a job opening.
* Allow applicant to upload resumes.
* Shortlist candidates viable for the job.

### Non-Functional Requirements

* Speed of the system.
* Security of the transactions.
* Reliability if the system.
* Response time.

## Preliminary Data Processing and analysis.

From the observations and the records reviewed the conclusion arrived at was that there is a great need for a machine learning model especially in Kenya as there is no model available here. It was also clear that the model will reduce workload and save time.

## REFERENCES

Kumba Sennaar Machine Learning for Recruiting and Hiring – 6 Current Applications, May 2019.

Jobbatical 2020 Hiring statistics, trends and data: The ultimate list of recruitment styles, February 2020