**BINDURA UNIVERSITY OF SCIENCE EDUCATION**

**FACULTY OF SCIENCE AND ENGINEERING**

**COMPUTER SCIENCE DEPARTMENT**



**TUITION FEES CROWD-FUNDING SYSTEM USING ARTIFICIAL INTELLIGENCE TO CLASSIFY STUDENTS IN NEED OF FINANCIAL ASSISTANCE**

By

Supervised by

**(Computer Science department)**

*Project Submitted in partial fulfilment of the requirements of the Bachelor of Science (BSc) Honours*

*Degree in Computer Science*

December 2020

# APPROVAL FORM

The undersigned certify that they have supervised the student dissertation entitled “**TUITION FEES CROWD-FUNDING SYSTEM USING ARTIFICIAL INTELLIGENCE TO CLASSIFY STUDENTS IN NEED OF FINANCIAL ASSISTANCE”**

submitted in Partial fulfilment of the requirements for the Bachelor of Computer Science Honors Degree of Bindura University of Science Education.

……………………………. …………………………... ………/………/………..

STUDENT SIGNATURE DATE

……………………………... ……………………………. ……/………/………

SUPERVISOR SIGNATURE DATE

……………………………. …………………………….. ………/………/…………

CHAIRPERSON DATE SIGNATURE DATE

…………………………….. …………………………….. ………/………/………..

EXTERNAL EXAMINER SIGNATURE DATE

# ABSTRACT

This paper aims to shed light on how Crowdfunding methods and artificial intelligence can also be used in the tertiary education system to assist students financially in a more efficient manner. Web-based crowdfunding methods have managed to transform how businesses and foundations raise funds in this era. Today, with the fast development of the computer science and IT, especially the Internet, it is certain that the crowd is connected to the Web. Taking advantage of the crowd that is available online, crowdfunding platforms can be built to try and solicit funds from the group for a cause. The author first reviews the emergent literature on crowdfunding to identify and report on its diverse nature. The author at first had to look at other applications of artificial intelligence in crowd-funding systems that have been implemented in the past. The outcomes of the paper might be useful for passionate students who do not have enough financial support to be able to attain funds in a much convenient way and also further scholars can design future research strategies given the different angles of the phenomenon.

# DEDICATION

This research is dedicated to my brilliant, outrageously loving and supportive parents, not forgetting my always encouraging and ever faithful relatives. I am forever indebted to your love. I thank you for all the support and believing in my dream, may the good Lord continue to bless you all.

# ACKNOWLEDGEMENT

I am vastly grateful for the awesome support and whole hearted contribution rendered by my supervisor Mr. P. Chaka for his guidance, assistance and expertise. I also appreciate my beloved parents, for the love, care, moral support, and financial support. Above all, to the loving heavenly Father, whose eternal love has buoyed me up whereby His blessing kept me burning and moving, Praise be to Him. A special thank you goes to my classmates for the enormous effort they put in towards information sharing and for making this journey worthwhile, without them the past four years would have been far less rich, thank you.

Table of Contents

[APPROVAL FORM i](#_Toc62107978)

[ABSTRACT ii](#_Toc62107979)

[DEDICATION iii](#_Toc62107980)

[ACKNOWLEDGEMENT iii](#_Toc62107981)

[LIST OF FIGURES vi](#_Toc62107982)

[CHAPTER ONE 1](#_Toc62107983)

[1.1 introduction 1](#_Toc62107984)

[1.2 background 1](#_Toc62107985)

[1.3 Problem Statement 3](#_Toc62107986)

[1.4 Aims 3](#_Toc62107987)

[1.5 Objectives 3](#_Toc62107988)

[1.6 Research Questions 3](#_Toc62107989)

[1.7 Scope 4](#_Toc62107990)

[1.8 Justification of the study 4](#_Toc62107991)

[1.9 Assumptions 4](#_Toc62107992)

[1.10 Limitations of the study 4](#_Toc62107993)

[1.11 Definition of Key Variables 4](#_Toc62107994)

[1.12 Conclusion 5](#_Toc62107995)

[CHAPTER TWO: LITERATURE REVIEW 6](#_Toc62107996)

[2.1 Introduction 6](#_Toc62107997)

[2.2 Crowdfunding 6](#_Toc62107998)

[2.2 Other Applications of Artificial Intelligence in Crowd funding 7](#_Toc62107999)

[2.2.1 Success Prediction on Crowdfunding with Multimodal Deep Learning 7](#_Toc62108000)

[2.2.2 Topic Predictions and Optimized Recommendation Mechanism Based on Integrated Topic Modelling and Deep Neural Networks in Crowdfunding Platforms 9](#_Toc62108001)

[2.2.3 Predicting outcomes in crowdfunding campaigns with textual, visual, and linguistic signals 13](#_Toc62108002)

[2.3 Conclusion 20](#_Toc62108003)

[CHAPTER THREE: METHODOLOGY 21](#_Toc62108004)

[3.1 Introduction 21](#_Toc62108005)

[3.2 Reviewing Relevant Literature 21](#_Toc62108006)

[3.3 Research Designs 21](#_Toc62108007)

[3.4 Population and Sampling frame 22](#_Toc62108008)

[3.5 Experimental Design 22](#_Toc62108009)

[3.6 System Requirements 23](#_Toc62108010)

[3.6.1 Functional Requirements 23](#_Toc62108011)

[3.6.2 Non-Functional Requirements 23](#_Toc62108012)

[3.6.3 Tools 23](#_Toc62108013)

[3.7 System Development Methodology 24](#_Toc62108014)

[3.7.1 Software Prototype Model 24](#_Toc62108015)

[3.8 System Designs 25](#_Toc62108016)

[3.8.1 Entity-Relationship Model of the system 26](#_Toc62108017)

[3.8.2 Data flow diagram 27](#_Toc62108018)

[3.8.3 Description of what happens in a tuition fees crowd-funding system using artificial intelligence 29](#_Toc62108019)

[3.9 Conclusion 31](#_Toc62108020)

[CHAPTER FOUR: DATA PRESENTATION 32](#_Toc62108021)

[4.1 Introduction 32](#_Toc62108022)

[4.2 System testing 32](#_Toc62108023)

[4.2.1 The Hardware Used for Testing 33](#_Toc62108024)

[4.2.2 Conditions for testing 33](#_Toc62108025)

[4.3 Presentation of results 33](#_Toc62108026)

[4.3.1 a summary table of selected attributes of dataset 33](#_Toc62108027)

[4.3.2 Comparing Error Rate with the K Value 34](#_Toc62108028)

[4.3.3 Performance Testing 35](#_Toc62108029)

[4.4 Summary of results 38](#_Toc62108030)

[4.5 Conclusion 39](#_Toc62108031)

[CHAPTER FIVE: CONCLUSION AND RECOMMENDATION 40](#_Toc62108032)

[5.1 Introduction 40](#_Toc62108033)

[5.2 The reviewing of the Aims and Objectives 40](#_Toc62108034)

[5.3 Challenges faced 41](#_Toc62108035)

[5.4 Future Work 41](#_Toc62108036)

[5.5 Recommendations 41](#_Toc62108037)

[References 42](#_Toc62108038)

LIST OF FIGURES

[Figure 1Crowdfunding 6](#_Toc62107946)

[Figure 2Latent Dirichlet Allocation (LDA) block diagram 11](#_Toc62107947)

[Figure 3Basic deep neural network architecture 12](#_Toc62107948)

[Figure 4 Distribution memory paragraph vector model (PV-DM) and Distributed bag-of-words paragraph vector model (PV-DBOW) 18](#_Toc62107949)

[Figure 5 K-Nearest Neighbour Algorithm 22](#_Toc62107950)

[Figure 6 the software Prototyping Model 24](#_Toc62107951)

[Figure 7 System er Diagram 26](#_Toc62107952)

[Figure 8 Data flow diagram 27](#_Toc62107953)

[Figure 9 Use case Diagram 28](#_Toc62107954)

[Figure 10 overview tuition fees crowd-funding system using artificial intelligence 29](#_Toc62107955)

[Figure 11 Homepage 30](#_Toc62107956)

[Figure 12 Student log in page 30](#_Toc62107957)

[Figure 13 Student Profiles Page 31](#_Toc62107958)

[Figure 14 summary table of selected attributes of dataset 33](#_Toc62107959)

[Figure 15 value range 34](#_Toc62107960)

[Figure 16 False acceptance rate 35](#_Toc62107961)

[Figure 17 False Acceptance Vs Correctly classified 36](#_Toc62107962)

[Figure 18 False Rejection Rate 36](#_Toc62107963)

[Figure 19 False Rejection vs Correctly Classified 37](#_Toc62107964)

[Figure 20 Computational Time 37](#_Toc62107965)

[Figure 21 KNN classification Results 38](#_Toc62107966)

# CHAPTER ONE

1.1 introduction

Financial aid systems from charity organisations such as the higher life foundation from Econet have been around over the past years. These organisations have helped so many students through financial assistance to allow them to have access to education and improve their livelihoods. The distribution of these funds has seemed to be less accessible to everyone and also the old system present has restricted single individuals that do not belong to any organisation to also participate in giving financial assistance. Students are dropping out of tertiary yearly because of failure to pay fees due to economic hardships whilst there are well-wishers on the other hand willing to give a helping hand to help those students that are less privileged. And with all this going on, the only thing that is missing is a platform that is convenient and transparent to allow those willing to assist to give a helping hand. This platform eliminates the middle man and thus in a way improves transparency and convenience unlike current scholarships being offered which are processed over a long period of time and also the issue of nepotism within organisations offering these scholarships. The student proposes a system for Tuition fees crowd-funding system using Artificial intelligence to classify students in need of financial assistance at his tertiary institution. If properly done, this can benefit the students and more importantly the economy.

1.2 background

Globally, education is viewed as a powerful tool for the development of any society and the much-needed human capital (Thomas, 2015). In that direction, the primary goal of education to any given nation is to prepare individuals for the job markets by transmitting knowledge, skills, attitudes and cultural norms from the adult world to the younger one. Having an educated society comes with many benefits which include economic benefits as a result of having a multiskilled society which can help in improving all the sectors of our economy and also civic benefits whereby people with good education and gainful employment often give back to the community and they are more likely to become involved in volunteer work.

Other benefits includepersonal development whereby people with careers tend to lead more structured lives and have a stronger sense of responsibility, traits that serve as strength-builders in other areas of life. Education systems worldwide, Zimbabwe included are challenged by the crisis of high dropouts and repetition rates as a result of financial hardships (UNESCO, 2009). Use of a tuition crowd funding system, will be extremely beneficial as it ensures that students can be drew closer to financial aid like they have never been before.

1.3 Problem Statement

The Zimbabwean economy is not in a good shape currently and even though showing signs of recovery, the process might take time. Tertiary fees are now becoming a major challenge for many students around the country and many are dropping out. A huge percentage of Zimbabweans are facing financial challenges and some families that have been hit hard by the economic crisis have resorted to minimizing their expenses by cutting educational expenses. Scholarships are not readily available for grab to all students. Currently available scholarship programs available seem to have bottlenecks when it comes to their distribution. What is lacking is a centralized platform that is accessible to everyone and everywhere for offering financial assistance to students.

1.4 Aims

The main aim of this research is to try and make sure that students who are struggling to complete their studies at university as a result of financial causes can have access to financial aid through a crowd funding system. This research aims to make sure that those students and are truly in need of financial assistance can benefit from the system by using an artificially intelligence classification model that can classify students according to their financial background. The student’s aim is to develop a system that eradicates the barriers to access of financial aid.

1.5 Objectives

I suggest the following objectives:

1. To design and implement a crowd funding system for financial aid in tertiary institutions using artificial intelligence.
2. To assess the effectiveness of using Artificial intelligence in crowd funding for provision of financial aid in tertiary institutions.

1.6 Research Questions

1. How can we reduce the number of student dropouts as a result of financial difficulties?
2. How can we ensure that the right students are benefitting from the system?

1.7 Scope

The scope of this project is as follows:

1. Research on how crowd-funding can be applied to tertiary institutions.
2. Research and implement a classification algorithm suitable for student profiling into classes in relation to their financial backgrounds.

1.8 Justification of the study

Having a tuition fees crowd funding system helps to reduce the numbers of student dropping out of university due to increased accessibility and also the incorporation of the Artificial Intelligence increases the accuracy to ensure that the correct people benefit from the system.

1.9 Assumptions

The research made some assumptions as the basis of study. The first assumption was that we could obtain a sample dataset from the university for training our classification model that fits the criteria of the students. The second assumption was that there were no external factors (political, economic or social) locally and globally that had a direct effect on people’s willingness to give to charity.

1.10 Limitations of the study

The study being undertaken by the student requires use of secondary data that has been collected for other purposes to be able to come up with a successful predictive model. Primary data is difficulty to use because it is almost impossible to collect accurate data about student finances through interviews.

1.11 Definition of Key Variables

Tertiary Institution – All universities and polytechnic colleges within the country

Scholar – Any academic who is studying or researching on a particular topic in any field of study

Classification – systematic arrangement in groups or categories according to established criteria

1.12 Conclusion

Zimbabwe among other nations is fighting to achieve the United Nations goal of education for all which is a means in achieving sustainable development. The Nziramasanga Commission (1999: 172) state that countries throughout the world are working to ensure education for all and Zimbabwe has the same goal. Ease of access to financial aid is one way we can use to move towards achieving this goal. It is for this reason that this project is important. In short, this chapter focused on problem identification and the next chapter is on literature review which focuses on other researches relevant to this study that have been written in the past years.

# CHAPTER TWO: LITERATURE REVIEW

2.1 Introduction

The main aim of this chapter is to look at other applications of artificial intelligence in crowd funding systems implemented in the past. It will closely look at issues highlighted in these research papers that include shortcomings and advantages surrounding the technology. There will be a few suggestions given to try and ensure the algorithm is efficient in its use.

2.2 Crowdfunding

Crowdfunding is any financing method that involves taking small amounts of money from a large number of individuals. The people who fund these projects and entities may do so without expecting anything in return, they're donations to a cause they support. Others fund these projects in exchange for products, services, or equity in the entity. Crowdfunding combines the concepts of crowdsourcing and microfinancing, bringing together various individuals who commit small amounts of money to projects, causes or entities they want to support. This is often done through websites that make it easy for entities to find potential funders. Many authors have researched on the topic of Crowdfunding and their subsequent research has highlighted several issues and shortcomings surrounding the technologies, as well as how helpful it is if implemented in the right way.

According to Miller et al (2018) web-based crowdfunding methods have transformed the ways businesses and foundations raise funds. One survey found that U.S. businesses raised a total of $1.04 billion through crowdfunding in 2018.

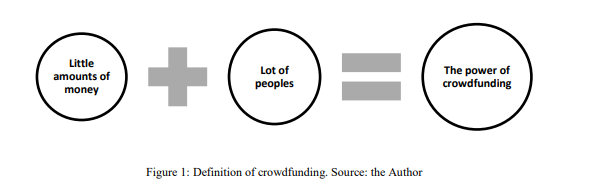


Figure Crowdfunding

The crowd can be defined as a multitude of individuals gathered in one place and in this case, the place is Internet. Today, with the fast development of the computer science and IT, especially the Internet, it is certain that the crowd is connected to the Web. O'Reilly (2007) defined the Web 2.0 as “a primarily collaborative web platform where individuals can share their resources” and thanks to that, the crowdfunding can be developed. For Kleeman et al (2008) the Web 2.0 is a critical instrument that has facilitated the access of the “crowd”.

2.2 Other Applications of Artificial Intelligence in Crowd funding

### 2.2.1 Success Prediction on Crowdfunding with Multimodal Deep Learning

The booming trend of crowdfunding has drawn much attention from academia. The conventional approach was to build a machine learning classiﬁer like SVM based on Meta features from campaign proﬁle. Greenberg et al. (2013) showed an improvement by utilizing various decision tree algorithms and SVM trained with features such as whether the video was present, the sentence count in proﬁle, project goals, project duration, and other possible additional factors like creators’ demographic attributes. Some advanced approaches utilize textual description while certain models additionally exploit dynamic information by monitoring social media or crowdfunding campaign. Mitra and Gilbert (2014) analysed the linguistic features with 59 other common features to predict project success.

Yuan et al. (2016) proposed a text analytic topic framework to predict the fundraising success by extracting latent semantics from the text description with a combination of common numerical features. Etter et al. (2013) and Zhao et al. (2017) studied dynamic time-series factors by tracking the social media and monitoring the dynamic features like backer and money pledged status during the campaign. Zheng et al. (2014) found the degree of the creator’s social network is positively associated with project success since creators can broadcast their crowdfunding projects to more audiences through the social network. Li et al. (2016) formulated the success prediction problem from the aspect of censored regression and achieved better performance utilizing temporal features from pledging and social media dynamics. From the above descriptions, we can observe that most of the previous works focused on textual proﬁle and post launch information which makes both project creators and platforms not able to predict the outcome in a timely manner. Therefore, to make pre-posting prediction possible, our approach focuses on the joint analysis of the textual and visual information collected from the pre-launch stage, which has not been fully explored yet in previous studies.

1. **Multimodal Analysis**

The multimodal approaches using text and image joint analysis have been explored in social media quite a while, e.g., multimodal semantic analysis (Roller and Im Walde, 2013), multimedia market evolution monitor (Zhang et al., 201), and multimodal news analysis (Ramisa Ayats, 2017). To encode visual information, most of the earlier approaches relied on hand-crafted features combined with methods to aggregate manually engineered descriptors before the rise of (CNN) convolutional neural networks. The Bag-of-Visual-Words (BoVW) model was the common choice of image feature representation. It would collect code words from feature descriptor like Scale-Invariant Feature Transform (SIFT), and then learn the codebooks from unsupervised learning.

In more recent works, hybrid architecture was introduced to leverage the best of the BoVW and deep neural networks. One approach was to combine the pre-trained deep features with BoVW, while some other approaches projected the hand-crafted feature descriptors to lower dimensionality and feed them to neural networks. The pre-trained CNN in large databases, like ImageNet, can be used as off-the-shelf feature extractors. The transfer feature learning from existing deep convolutional neural networks showed promising results in varied researches (Wangetal 2016). Nevertheless, this approach has not been proved to be an effective corollary in crowdfunding project success prediction. Moreover, the adoption of existing approach was hindered by the heavier sparsity of image fusion which need to be carefully addressed in this problem.

### 2.2.2 Topic Predictions and Optimized Recommendation Mechanism Based on Integrated Topic Modelling and Deep Neural Networks in Crowdfunding Platforms

The accelerated growth rate of internet users and its applications, primarily e-business, has accustomed people to write their comments and reviews about the product they received. These reviews are remarkably competent to shape customers’ decisions. However, in crowdfunding, where investors finance innovative ideas in exchange for some rewards or products, the comments of investors are often ignored. These comments can play a markedly significant role in helping crowdfunding platforms to battle against the bitter challenge of fraudulent activities contributions (Mollick 2013; Ahlers et al. 2015). Taking advantage of language modelling techniques and aiming to merge them with neural networks to identify some hidden discussion patterns in the comments.

The objective was to design a language modelling based neural network architecture, where Recurrent Neural Networks (RNN) Long Short-Term Memory (LSTM) is used to predict discussion trends, i.e. either towards scam or non-scam. LSTM layers are fed with latent topic distribution learned from the pre-trained Latent Dirichlet Allocation (LDA) model Zheng et al. (2014). In order to optimize the recommendations, Particle Swarm Optimization (PSO) was used as a base line algorithm. This would help investors find secure projects to invest in (with the highest chances of delivery) within their preferred categories. Prediction accuracy, an optimal number of identified topics, and the number of epochs, were used as metrics of performance evaluation for the proposed approach. Results were compared with simple Neural Networks (NNs) and NN-LDA based on these performance metrics. The strengths of both integrated models suggest that the proposed model can play a substantial role in a better understanding of crowdfunding comments.

1. **Predictions and Recommendations in Crowdfunding**

There are various studies on crowdfunding that target to predict diﬀerent trends and project success. A tool is built to get reviews on their project ideas for a start-up. Some studies have explored diﬀerent linguistic features, speciﬁc patterns, and writing styles of project creators to reveal the impact of language on the success or failure of a campaign. Crowdfunding success prediction is estimated through a text analytics approach, where LDA is used to extract semantic features out of the text, along with feature selection, and data mining. In a similar work, crowdfunding updates are analysed by using LDA to classify the updates into diﬀerent topic categories.

1. **Topic Models and RNNs**

To capture the semantic features of the text, topic models play a vital role. The semantic features are extracted as latent topics. There have been numerous studies and applications of topic models since this idea was first introduced by Bleietal. These studies have covered many research areas ranging from scientific studies to mathematical equations. Recently, many studies are using deep learning methods in combination with topic models. RNN is used along with topic models to enhance the performance of topic modelling for scientific texts.

A neural topic model has been presented. Some recent studies have paid more attention and focused on neural variation inferences in order to train the topic models. Moreover, for sequences that have long term dependencies, RNN proved to be an effective solution (Wangetal 2016). Other studies that have used RNNs to model different language-related problems include handwriting recognition LATEX modelling, and semantic parsing, etc. This model is driven from a hybrid of the approaches which are based on a joint topic language model. The motivation behind these architectures is to extract the latent topics through topic models, e.g., LDA and these pre-trained models are used for deep neural networks, such as RNNs or CNNs, modelling. In other studies, both topic models and neural networks are trained together, which we also aim to perform. A recent work has proposed a Sentence Level Recurrent Topic Model (SLRTM), where for each sentence, a topic is decided based on a non-sequential Dirichlet structure similar to LDA.

Therefore, it is not much useful for capturing long-range temporal dependencies, also as it uses the whole vocabulary, which enriches it with lots of features. In this case, only topics are used, not the entire dictionary. There are other models introduced, which are based on recurrent latent variables; in these models, RNN is enabled with latent variables in order to cater to the inconsistencies in the input data.

Though, these models focus on images and speech data; while we use a discrete input data space and uses text data and numeric data. In the previous studies, LDA and deep learning models are combined to either improve the language modelling or on the same word text. In this case, a recommendation model is built on top of LDA-LSTM hybrid model. User preferences are added as an extra layer of input. One key diﬀerence of our model is that the latent topics are used in the LSTM layer while other models, such as and, integrate them in the output layer of LSTM

1. **Language Modelling**

Language modelling depends on a function learning that calculates the log probability of an activity as log(ω|model), or a sentence as ω = (ω1, ω2,…, ωn). This function is then used for the prediction of the next word or activity. It can also be used in different other ways, e.g., LDA uses a bag-of-words approach. Alternatively, it can be used in RNN modelling to prevent temporal information loss, to model log(ωn| ω1, ω2,…, ωn-1, model).

Latent Dirichlet Allocation (LDA) in this proposed system, LDA is used for topic modelling, Figure below presents the basic block diagram of LDA, where a user comment is treated as a document and fed to the LDA model, after some pre-processing. LDA results into clusters of similar words indicating a topic or theme. This process is repeated for all the comments in the dataset (Allison et al. 2017).

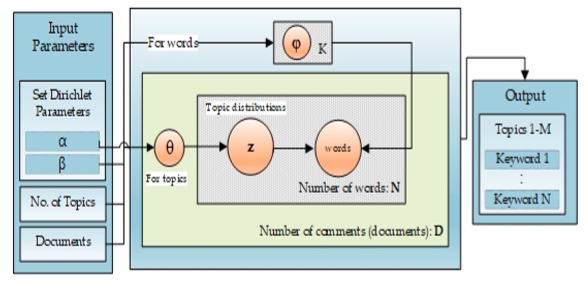


Figure Latent Dirichlet Allocation (LDA) block diagram

**Long Short-Term Memory (LSTM)**

The traditional topic modelling approaches have some limitations when it comes to context learning. For any language model, temporal aspects are also fundamental. Therefore, a well suitable method is required to perform this task. An RNN type, LSTM, can effectively learn the context and temporal features and can better classify or predict, primarily when we have large data sets with time-series information (Greenberg et al. 2013). RNNs are the type of networks that generate recurrent connections to memorize. Language models based on recurrent neural networks have lately established state-of-the-art performance in different applications. The dynamic temporal behaviour of RNN makes them favourable for sequential classification-based problems. For training, it takes the first-word w from the sequence of input; the output h0 along with the next word w1 is taken as input in the next step, and so on. This way, it keeps remembering the context while training, as shown in the figure below.

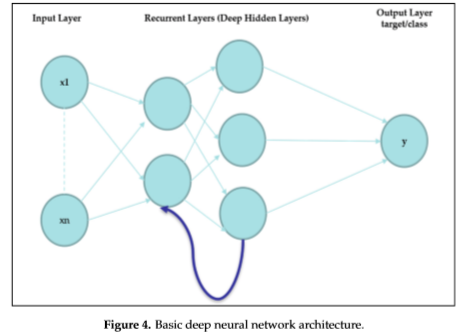


Figure Basic deep neural network architecture

Simply deﬁning an RNN, we can elaborate it in terms of (P, S, δ) where P represents the inputs, S represents the states, and δ is the transition function of the neural network. For example, a traditional language model based on RNN considers a document as a sequence. To predict the next word, an LSTM is trained which also takes into account the previous words. Hence, it maximize sp(wt|wt−1|,wt−2,... , w0; model). The input words are transformed to vectors in Σ which eventually is used for the LSTM state update. For the output, a projection of st is required into a vector of the size of the dictionary and is followed by an activation function, e.g., Softmax. However, challenges occur with more extensive size dictionaries.

1. **Particle Swarm Optimization (PSO)**

The PSO was first proposed by J. Kennedy and R. Eberhart in 1995. It is a population-based optimization algorithm, which became very famous because of its continuous optimization process towards the best solution. It is derived from the concepts of swarming habits of animals, e.g., fish or birds and also from genetic algorithms. At a given time t, PSO upholds multiple possible solutions, each represented by a particle. The fitness of these solutions is calculated during each iteration by using an optimization function. There is a certain velocity which each particle has to move with, in order to reach the maximum value, returned by the objective function. At each iteration, the particle’s position and velocity are updated according to the following Equations (1) and (2):

V (t + 1) = W \* V (t) + c1 \* r1 x [Y(t) − X(t) ] + c2 \* r2 × [G(t) − X(t)] (1)

X(t + 1) = X(t) + V (t + 1) (2)

where V is the particle’s velocity, X(t) is particle’s current position at time t. Y(t) is the individually best solution of the particle at time t, and G(t) is the global best solution of the swarm at time t. W is the coefficient of inertia, usually ranges between 0.8 to 1.2. r1 and r2 are the random numbers generated in the range [0,1], and c1 and c2 are the cognitive and social coefficients, respectively. These coefficients are also known as learning factors, and their value is usually kept as 2.

### 2.2.3 Predicting outcomes in crowdfunding campaigns with textual, visual, and linguistic signals

This research introduces a neural network and natural language processing approach to predict the outcome of crowdfunding start-up pitches using text, speech, and video metadata in 20,188 crowdfunding campaigns. The study emphasizes the need to understand crowdfunding from an investor’s perspective. Linguistic styles in crowdfunding campaigns that aim to trigger excitement or are aimed at inclusiveness are better predictors of campaign success than firm-level determinants.

Consequently, potential backers in crowdfunding are looking for potential cues to reduce uncertainty and predict new venture success when making their capital contributions (Mollick 2013; Ahlers et al. 2015). One way for innovators to overcome this uncertainty is to signal competence trust, arising from expectations about the competence of the innovator, to create a higher receptivity among potential contributors. Prior work has shown that impression management (Parhankangas and Ehrlich 2014), competence signalling (Gafni et al. 2019), and persuasion (Allison et al. 2017) may all affect crowdfunding positively. However, prior work has found mixed evidence on the role of visual and textual cues. While Parhankangas and Renko (2017) find that commercial entrepreneurs need to primarily focus on product, or firm and entrepreneur-related signals in their textual descriptions, other work shows that in low attention states visual cues work best, while textual information become only relevant if a high attention has been triggered previously (Allison et al. 2017). Hence, the effectiveness of a crowdfunding campaign pitch is inextricably linked to the various media involved.

 Therefore, computational techniques are explored to predict crowdfunding campaign success based on the informational cues provided within campaign text, speech, and videos. Advances in data processing and machine learning allow new ways of analysing data and may have profound implications for empirical testing of lightly studied, yet complex, empirical relationships. That being said, an idea is proposed that, new forms of internet-mediated capital, such as crowdfunding, provide comprehensive and potentially computable signals to predict outcomes or provide recommendations. For instance, crowdfunding could be considered as perhaps the biggest open laboratory to study the interaction of inventors and investors at large scale.

In this research, a novel method is proposed that combines neural networks and text-mining to identify features of successful crowdfunding projects, using transformed text, speech, and video content. Using text, speech, and video object–related meta-data in 20,188 crowdfunding campaigns, our analysis employs natural language processing techniques and neural network models to predict the success of crowdfunding campaigns. Based on word and paragraph vector models of text, speech, and video information, a feature-union model achieves a prediction accuracy of 73% in explaining campaign success or failure. Besides, derive dialectic particularities are derived in text, speech, and video characteristics that determine whether campaigns are more likely to be successful. The study emphasizes the need to understand crowdfunding from a consumer’s and future investor’s perspective. Linguistic styles in crowdfunding campaigns that aim to trigger excitement, or are aimed at inclusiveness, are better predictors of campaign success than firm-level determinants. At the contrary, higher uncertainty perceptions may substantially reduce evaluations of new products and reduce purchasing intentions among potential funders. Findings emphasize that positive psychological language is salient in environments where objective information is scarce and where investment preferences are taste based.

1. **Methodology used in this research**

This research payed attention to the textual and linguistic context of crowdfunding campaigns. Early work here focused on the prediction of campaign success using text-mining features from project descriptions (Greenberg et al. 2013). Researchers used decision tree (DT) algorithms and support vector machines (SVC) to train a machine learning classifier on explaining campaign success (Greenberg et al. 2013). Models achieved 68% accuracy with their respective datasets, an improvement of roughly 14% over the related baseline. More recent research focuses on the predictive power of project description content, specifically the words and phrases project creators use (Mitra and Gilbert 2014). In here, linguistic features extracted from project descriptions were combined with other campaign features to predict crowdfunding success. Tools such as Linguistic Inquiry and Word Count (LIWC) (Pennebaker et al. 2001; Tausczik and Pennebaker 2010) infer psychologically meaningful styles and social behaviours from unstructured text (Mitra and Gilbert 2014; Desai et al. 2015; Kaminski et al. 2017). Mitra and Gilbert (2014) conclude that the language used in the project has a surprisingly high predictive power, accounting for about 59% of the variance around successful funding.

1. **Elaboration Likelihood Model (ELM)**

Based on the theory of the Elaboration Likelihood Model (ELM) (Petty and Cacioppo 1986; Bhattacharjee and Sanford 2006), Du et al. (2015) study the influence of project descriptions on crowdfunding success. Using constructs such as argument quality (number of words, readability regarding the Gunning Fog Index, sentiment ratio) and source credibility (previous campaign track record), the model predicts funding success with an accuracy rate of about 71–73%. Using campaign description text data only, Lee et al. (2018) present work building upon sequence-to-sequence (seq2seq) deep neural network models with an average 76% prediction accuracy on the first day of project launch.

Lastly, other approaches focused on contextual variables such as the social network activity of campaigns to predict funding success. For instance, the size of the social network of founders positively influences project success (Mollick 2014b). Social media activity explains some 75% variations in campaign success when conditioning on early project stages (Lu et al. 2014). More recently, studies already began to employ machine learning classifiers to predict the temporal backing patterns using project-based information and social features, obtained from Twitter (Etter et al. 2013; Li et al. 2016; Tran et al. 2016) and backer-network graphs (Etter et al. 2013). Using a k-nearest neighbours (kNN) classifier and a Markov Chain, Etter et al. (2013) predicted the trajectories of money pledged to campaigns. Drawing upon a dataset of 16,042 campaigns, a logistic regression and linear SVC estimator reach an accuracy of more than 76% (a relative improvement of 4%), 4 hours after the launch of a campaign (Etter et al. 2013). A similar stream of research focuses on the social dimension of campaigns considers the sentiment from user comments on campaign updates (Desai et al. 2015; Lai et al. 2017). Results suggest that the text sentiment and quality in comments one week after launch are very predictive for a campaign’s outcome. As Greenberg et al. (2013), Hui et al. (2013), and Yuan et al. (2016) conclude, prediction models can be used to give feedback on proposed campaigns or as a tool to match projects with potential investors (An et al. 2014).

Notwithstanding these contributions, there is a dearth of studies considering actual speech content and visual campaign narratives to predict crowdfunding success. For instance, analysing the linguistic style of crowdfunding pitches enables to conclude about revealed emotions and speech characteristics of creators (Kim et al. 2016), to distinguish between social or commercial entrepreneurs (Parhankangas and Renko 2017), or to separate conventional from “lead users” (von Hippel 1986) induced crowdfunding campaigns (Kaminski et al. 2017; Oo et al. 2018).

Concerning the analysis of video content, only standard approaches have hitherto been used for the evaluation of qualitative content and to measure the subjective perception of crowdfunding videos. Analyses mainly relate to the storyline and social construction (Doyle et al. 2017), perceived innovativeness, passion, preparedness, video quality, product appeal, perceived effort (Koch and Cheng 2016; Chan and Parhankangas 2017; Dey et al. 2017), and lead user appearance (Kaminski et al. 2017; Oo et al. 2018).

1. **Speech content**

As for speech transcription, the *Google Cloud Speech REST API* was used. Using a custom-build Python script, all project video files are first transformed into mono channel \*.flac audio files with *ffmpeg*. Audio files are subsequently uploaded into the Google Cloud to enable asynchronous English speech recognition via API, as a long-running operation until the end of an audio file. As for the Speech API, file URLs are used as input, while the output returns the *transcript* text of the speech and the average confidence of this transcription in a Pandas Data Frame. The confidence value is an estimate ranging from 0 to 1, indicating how confident the Speech API is in a given transcription. A higher number indicates a greater likelihood that the recognized words are correctly transcribed. However, it cannot be guaranteed that they are correct.

There are a few situations where some audio file transcriptions indicate a low confidence level, close to 0. In those cases, for instance, the original video either does not contain any speech signal other than music, a different language, or infrequent, untrained words. Only transcriptions with a confidence score of 0.80 are considered. An inspection showed that the entire body of text is sufficient, although the software does not recognize new brand (project) names and sometimes has problems with sentences that contain long stylistic pauses. With regard to pre-processing, the exact same pre-processing techniques as outlined above is applied.

1. **Video object recognition**

For visual content, analysis of all Kickstarter video files was done with the *Google Cloud Video Intelligence REST API*. The goal is to detect all different objects and their duration of appearance in each streaming video file. The *analyse labels* function from the Google Cloud Video Intelligence API is used to source object labels and their duration of appearance (*shots*) in a video sequence. In total, the data comprises 922,678 identified objects in 18,810 total video minutes with an average runtime of 1:20 min. A manual inspection of a few videos and respective video tags shows that the API has indeed a high accuracy identifying objects and events in videos. For video tags, no application of additional text cleaning was necessary.

1. **Models**

At the core of the system is a combination of an unsupervised learning of multidimensional vector representations of words and documents, respectively, as well as a supervised labelling approach with regard to campaign outcomes. The very first challenge to process natural language using deep learning is to represent the textual data in the form of fixed-length numerical data as input for deep neural networks. The most common approaches are bag-of-words (BOW), n-grams, and one-hot vectors. However, such models either do not preserve the word order or generate the same representations for different ordered sentences with the same words. Mentioned methods maintain the short context but tend to lose the overall semantics and fail drastically, when the length of a sentence is too long (Mikolov et al. 2013b). Hence, for employing neural network language models, word and paragraph vectors are used, *Doc2Vec*, preserving the semantics of natural language information. These vectors are learnt using the models as discussed by Mikolov et al. (2013b) and Le and Mikolov (2014). Paragraph vectors are an extension to Word2Vec. While Word2Vec learns to project words into a latent N-dimensional space, Doc2Vec aims at projecting a document into a latent N-dimensional space. As such, Doc2Vec is used to learn fixed-length vector representations for each word and paragraph in a high-dimensional continuous space. More precisely, word vectors are used with a feed-forward neural network, using a bag-of-words and skip-gram approach as outlined by Le and Mikolov (2014).

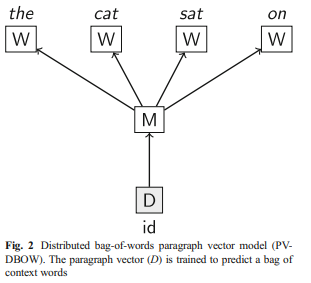
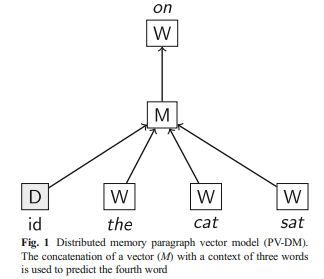


Figure Distribution memory paragraph vector model (PV-DM) and Distributed bag-of-words paragraph vector model (PV-DBOW)

Paragraph Vectors (PV) are embedding vectors which capture the overall semantic meaning of a text of variable length (“document to vector”). “The name Paragraph Vector is to emphasize the fact that the method can be applied to variable-length pieces of texts, anything from a phrase or sentence to a large document.” (Le and Mikolov 2014). Models of learning word vectors inspire the approach of learning paragraph vectors.

According to Mikolov et al. (2013b), models using large corpora and a high number of dimensions, like the PV-DM (skip-gram) model, promise a high accuracy, both on semantic and syntactic relationships. Performance benchmarks of the Paragraph Vector approach, in comparison to other approaches such as Recursive Neural Tensor Network (RNTN) (Socher et al. 2011), Naive-Bayes Support Vector Machine (NBSVM) (Wang and Manning 2012), or Restricted Boltzmann Machines model (RBM) with bag-of-words (Dahl et al. 2012), indicate a lower error rate (Le and Mikolov 2014).

1. **Classification**

After training the paragraph vectors, the 200-dimensional features are fed into several distinct classifiers. In total, six widely used parametric and non-parametric classifiers are being applied. As linear classifiers, (1) Logistic Regression and a (2) Linear Support Vector Classification (LinearSVC) are considered. (3) Gaussian Naive Bayes (GaussianNB), (4) Support Vector Classifier (SVC) are used as non-linear classifiers with a radial basis function kernel (rbf), the (5) XGBoost (XGBoost), which is a scalable tree boosting system (Chen and Guestrin 2016), and a (6) Multi-Layer Perceptron (Neural Network), which is a neural network model with 100 hidden layers and a rectified linear unit (ReLU) activation function (Nair and Hinton 2010). As it concerns the parameters of the classifiers, the classification model is trained with Grid-Search, supported fivefold cross-validation, and iterate over a comprehensive set of individual hyper parameters in scikit-learn (Pedregosa et al. 2011).

The final results represent the outcome of each best-selected classifier, by accuracy. Crowdfunding success is implemented as a binary variable indicating whether the campaign reached the funding goal (1) or not (0). This binary representation also resembles the “All-or-Nothing” (AON) approach of Kickstarter (Cumming et al. 2015). The AON model involves the entrepreneurial firm setting a fundraising goal and keeping nothing unless the goal is achieved. Each classifier is trained using the transformed paragraph vectors as the features (inputs) and labels as outputs.

2.3 Conclusion

In a nutshell, Artificial intelligence has proved to be a useful tool towards enhancement of crowd funding. Weaknesses have been clearly and explicitly discussed, with advantages and disadvantages of the existing implementations discussed. There is need to address several of these concerns in order to come up with a solution to the current problem under study, and these will be looked at during the course of development. The next chapter is on research methodology and it focuses on how the research under study was done.

# CHAPTER THREE: METHODOLOGY

3.1 Introduction

The orderly, theoretical analysis of the methods applied to a field is called methodology. The main purpose of this chapter is to identify how the methodology for the system design, data collection approaches, research design and how they were carried out in order to achieve a predictive model that is clear and makes reliable predictions and produce high level accuracy. The researcher clarifies how data for the research project was gathered in this chapter. It explains the procedures used to research. It also provides a detailed plan of which research designs where used and the data collection methods.

3.2 Reviewing Relevant Literature

The researcher reviewed the literature, this was looking on the following: the definition of crowd funding systems, crowd funding system techniques already implemented, the necessities of a crowd funding system, functional and security requirement of crowd funding systems, the implementation of crowd funding systems around the world and security risks in using internet for crowd funding.

3.3 Research Designs

This project also used applied research to create a specific artificial intelligence (AI) tool based on a model and tested its performance. In this research, an AI model was designed, based on the k-nearest algorithm using python. This was then developed into a working prototype for purposes of testing. The algorithm was subjected to both a training phase and a testing phase. The holdout method was chosen, where a portion of the available data was used for training, while the balance was used for confirmation of prediction accuracy. The performance of the tool was analysed based on its accuracy in classification of the data.

3.4 Population and Sampling frame

Population is a group of people from which samples are taken on which a research will be based. In this research, the researcher had to make use of an opensource dataset. A sample is therefore expected to mirror the population from which it comes although a sampling error can occur. There are two types of sampling errors which must be guarded against, and these are sampling bias, which is a tendency to favour the selection of units that have particular characteristics, and errors which are due to chance.

3.5 Experimental Design

The experiment was conducted using the K nearest algorithm on the data. In statistics, the k-nearest neighbours’ algorithm (k-NN) is a non-parametric method proposed by Thomas Cover used for classification and regression. But KNN is widely used for classification problems in machine learning. KNN works on a principle assuming every data point falling near to each other is falling in the same class. That means similar things are near to each other. In both cases, the input consists of the k closest training examples in the feature space. In k-NN classification, the output is a class membership.

An object is classified by a plurality vote of its neighbours, with the object being assigned to the class most common among its k nearest neighbours (k is a positive integer, typically small). If k = 1, then the object is simply assigned to the class of that single nearest neighbour. To get the right value of K, you should run the KNN algorithm several times with different values of K and select the one that has the least number of errors. The right K must be able to predict data that it hasn’t seen before accurately.

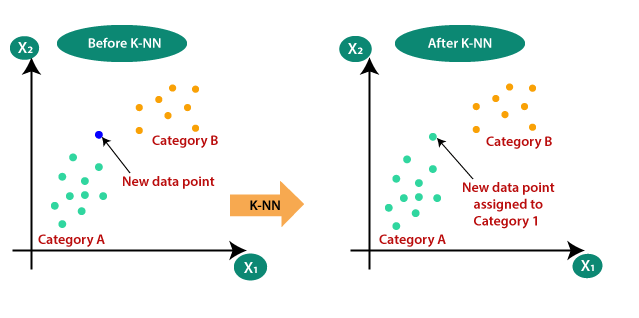


Figure K-Nearest Neighbour Algorithm

3.6 System Requirements

Technical specifics of the method used to get to the resolution and the implementation is described in the system design. It was essential to design a crowd funding system that accurately categorize students according to their financial backgrounds and allows students to virtually attain funds for their tertiary studies and allow them to pursue their dreams as mentioned in chapter one.

### 3.6.1 Functional Requirements

Services which the system should provide and how the system should behave in particular situations, the system should be able to:

* Categorise students accordingly.
* Accommodate interaction of students and donors.
* Ensure secure transfer of funds from donors.

### 3.6.2 Non-Functional Requirements

**Performance requirements**: the system must process transaction efficiently in term of speed and device resources

**Accessibility requirements**: Authenticated student must be able to access the system.

**Availability requirements**: the system services must be available to every authenticated user every day.

### 3.6.3 Tools

Tools used to develop the proposed system:

**Software**

* Programming Language-python for algorithm development
* HTML and CSS5 for user interfaces
* JavaScript for front end scripting and PHP for server-side scripting
* MySQL relational database
* Scikit-Learn, Pandas, numpy
* IDEs Used- Visual Studio 2019

**Hardware**

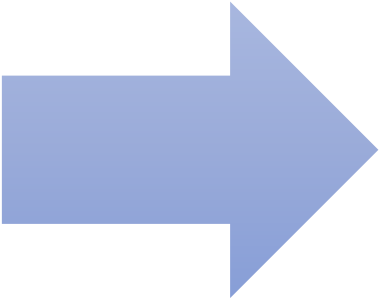
* Processor-Intel Core i7
* RAM-8GB
* HDD-1TB

3.7 System Development Methodology

A system development methodology refers to the framework that is used to structure, plan, and control the process of developing an information system. A wide variety of such frameworks have evolved over the years and each with its own recognized strengths and weaknesses. One system development methodology is not necessarily suitable for use by all projects and this proposed system was designed using the Software Prototype Model.

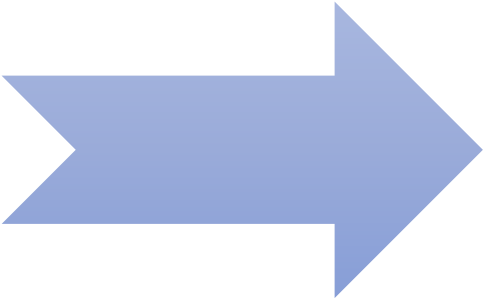
### 3.7.1 Software Prototype Model

Software prototyping is becoming very popular as a software development model, as it enables to understand customer requirements at an early stage of development. It helps get valuable feedback from the customer and helps software designers and developers understand about what exactly is expected from the product under development.



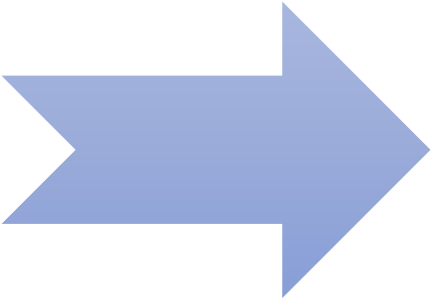
**Requirement**

**gathering**



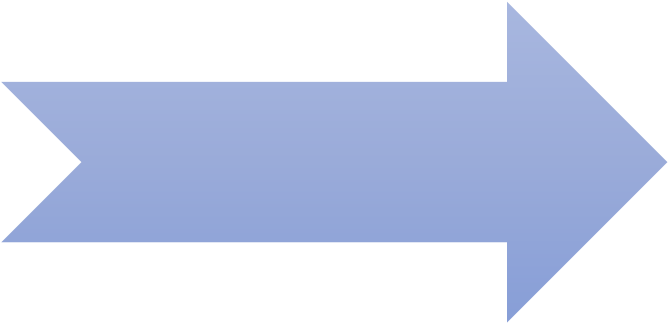
**Initial**

**Prototyping**



**Prototype**

**Review**



**Revise and enhance**

**prototype**

Figure the software Prototyping Model

**Requirement Identification and gathering -** This involves understanding the basics product requirements. This means that all other aspects like performance and security can be ignored at this stage.

**Design and developing the initial Prototype -** In this stage the initial prototype is developed, where the very basic requirements are showcased. These features may not work exactly in the same manner internally in the final software developed.

**Review of the Prototype -** The prototype developed is then presented to the customer and the other important stakeholders in the project. The feedback is then collected in an organized manner and used for further enhancements in the product under development.

**Revise and Enhance the Prototype -** The feedback and the review comments from the end user are taken into considerations and some negotiations with the customer based on time and budget constraints and technical feasibility of the actual implementation. When all the requirements are met then the system is delivered to the user otherwise prototyping will continue.

Software prototype model has got many types and the researcher has adopted **Incremental Prototyping**. Incremental prototyping refers to building multiple functional prototypes of the various sub-systems and then integrating all the available prototypes to form a complete system.

3.8 System Designs

The system consists of a web application which is used as an interface between the system and both students and the donors. It also provides a platform that allows donor to make payments to the student’s account directly.

The student creates an account on the platform and logs in to view and update his or her profile. The Donor also has a profile where transaction logs and other account activities can also be monitored.

### 3.8.1 Entity-Relationship Model of the system

Entity Relational Diagrams was used to design the database of the prototype, the data modelling technique was used to create a graphical illustration of the entities, and the relationships between entities, within the system.

|  |
| --- |
| Donor Account  create  Student  Student Account  Donor  create  contacts  selects  has  Payment Method  has  mailbox |

Figure System er Diagram

### 3.8.2 Data flow diagram

A data-flow diagram (DFD) is a graphical representation of the flow of data through an information system. DFDs can also be used for the visualization of data processing (structured design). Data flow diagrams illustrate how data is processed by a system in terms of inputs and outputs. Figure below show data flow diagram of the proposed system.

|  |
| --- |
| dataset  Register Student account  Classification of student data  Results (Good, Average, Poor)  Training of a classification model  Complete profile  Analysis of data |

Figure Data flow diagram

3.6.3 Use case diagram

The use case diagram (Fig 16) below shows the interactions between the system and both student and the well-wishers or donors, which is all functionality a student and the donors can get from the system.

|  |
| --- |
| Student  Donor  Payment processor  Admin |

Figure Use case Diagram

### 3.8.3 Description of what happens in a tuition fees crowd-funding system using artificial intelligence

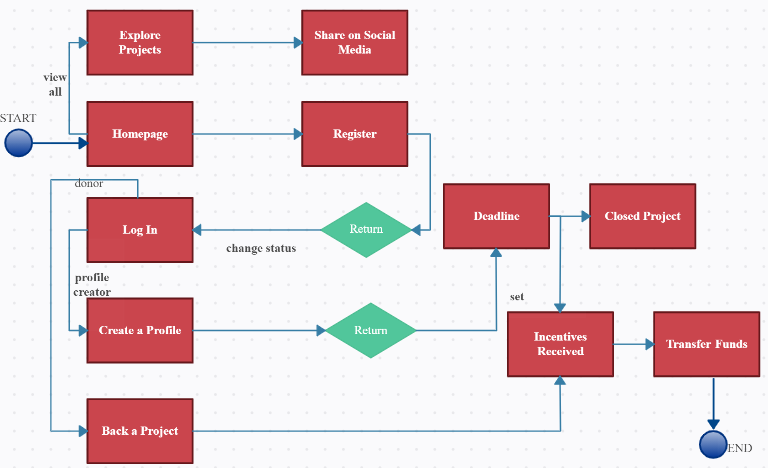


Figure overview tuition fees crowd-funding system using artificial intelligence

The proposed system above consists of a web application which is the main interface to the system. Both the students and the Donors have to create accounts to be able to interact with system. After logging in the student data is collected from the student profile which he would have completed and also from other secondary sources such as the university repository. On the server-side student data collected is fed to an already trained KNN classification algorithm running on the server for classification. Basing on the training model the algorithm will determine whether the student financial background is either good, average or poor. The classification process will allocate the student profiles priority values depending with the category which they are classified into. Profiles in the category labelled poor will have a much larger priority value compared to those in category average and category good. This prioritisation will work as a screening method to determine which profiles will appear first in line on the donation page.

**Screenshots**

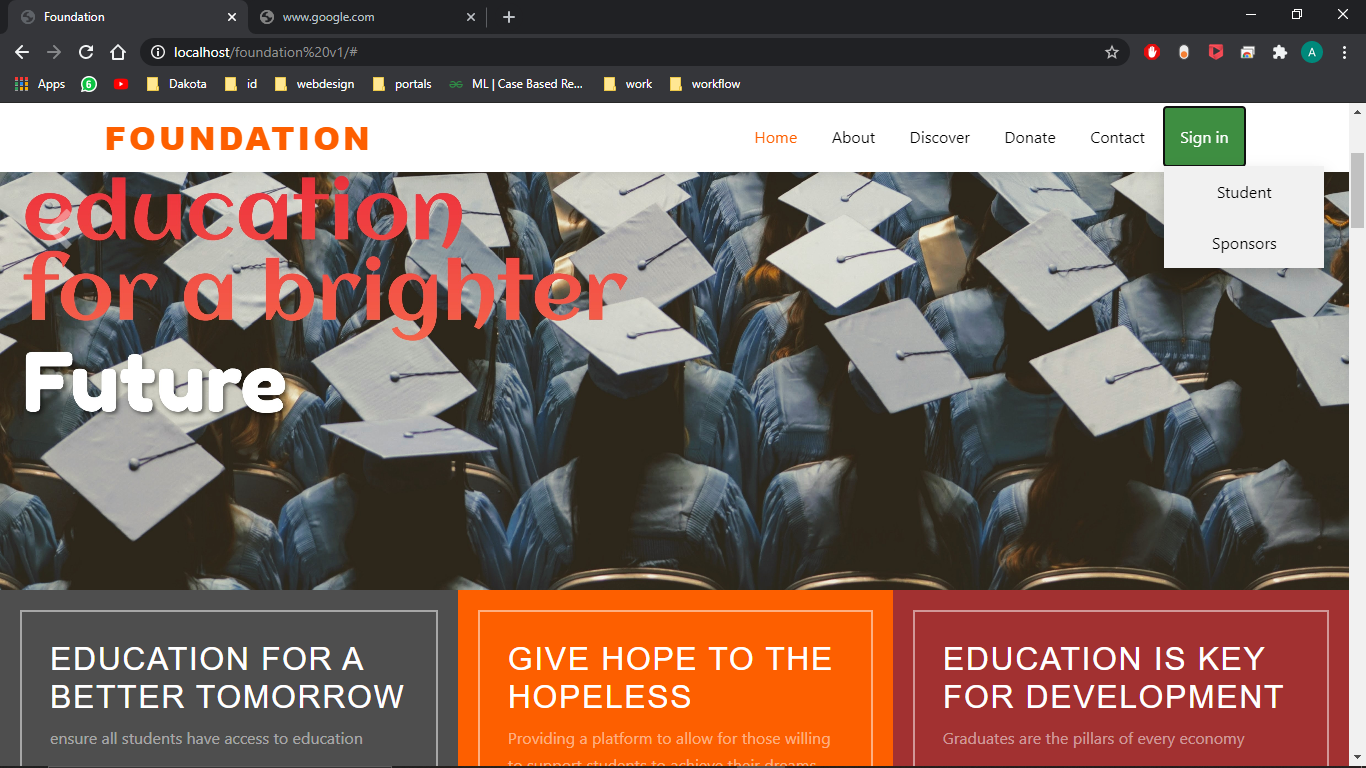


Figure Homepage

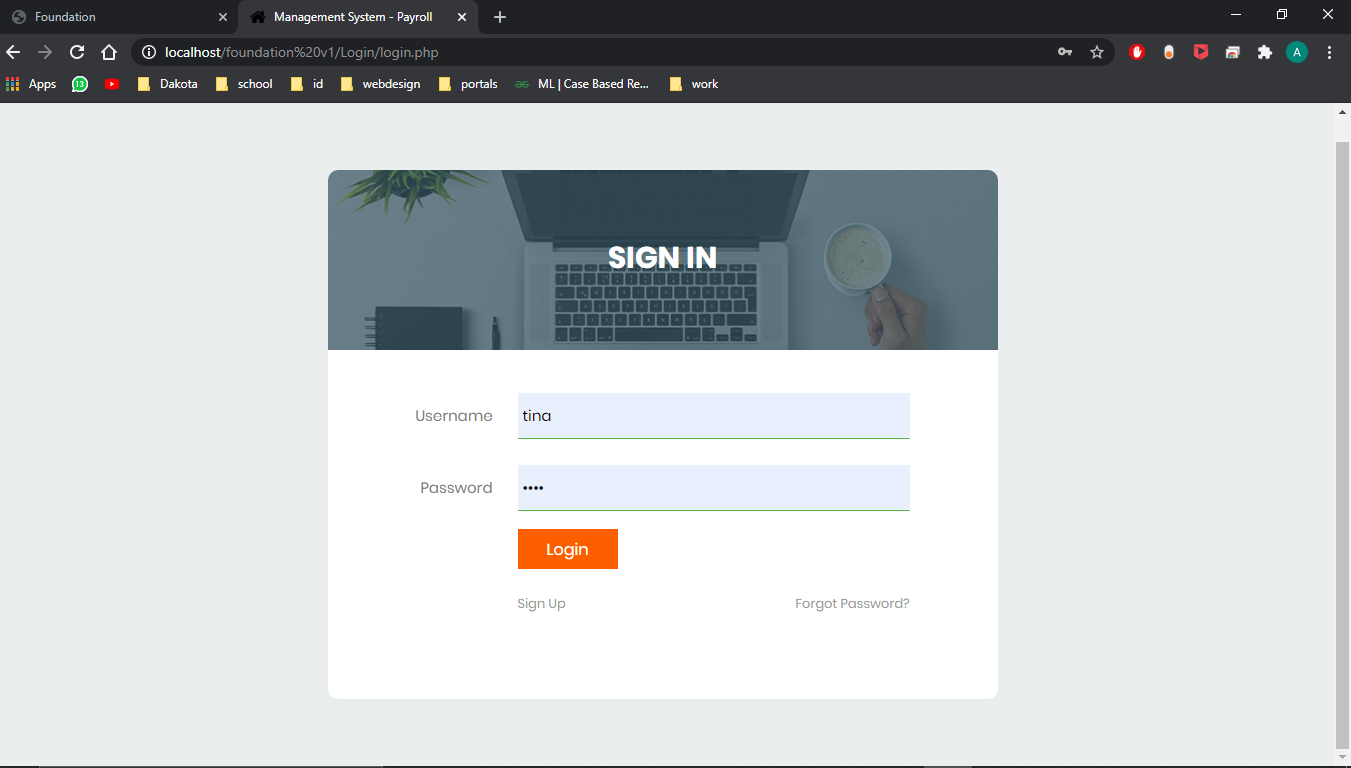


Figure Student log in page

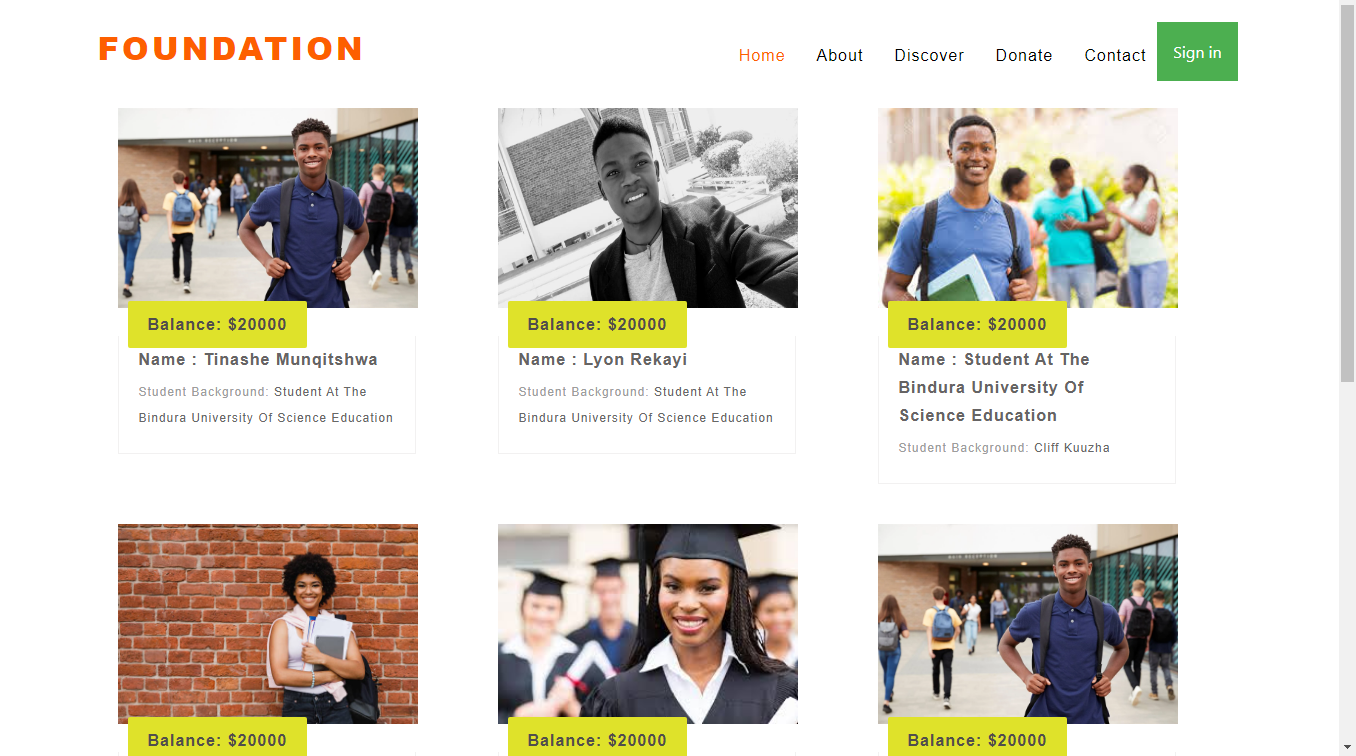


Figure Student Profiles Page

3.9 Conclusion

In a nutshell, this chapter dealt with the research methodology adopted during the study and it was based on issues to do with the research design, software design, system design process, development methodology and the implementation design. It also includes the sources of data, sampling procedures and presentation of data, data collection, as well as the validity and reliability of these instruments. The next chapter is on data presentation and analysis.

# CHAPTER FOUR: DATA PRESENTATION

4.1 Introduction

This chapter takes a look at the evaluation of the crowd funding System through various system tests and the analysis of effectiveness. The usability of the KNN classification model in its environment was evaluated based on the following software attributes which are scalability, stability, speed and compatibility. Different test to evaluate the performance of the system were carried out. The system’s behaviour response to comparison and classification of student data were analysed to give an evaluation of the system based on those attributes.

4.2 System testing

The system performance tests were carried out for the purpose of testing whether the system would be able to cope with the different variant environmental factors that may make it redundant. Known tests in software engineering under the field of system performance testing were used for the purpose of determining how the system will perform in terms of responsiveness and stability under a particular workload.

The researcher used performance testing for the measuring of other quality attributes of the system such as query response time, effectiveness in classifying students according to their true financial backgrounds and resource usage. Which strives to build performance into the implementation, design and architecture of a system.

For this research, the most common performance concerns related to a web application were adopted and these are the speed of query processing, the ability to handle stress thus concurrently handling queries efficiently and the ability to perform within capacity. In this case the researcher adopted black box testing which is a software testing method which tests the functionalities of the system without testing the internal functions and structures, it only tests the system as a whole for functionalities to check if it functions in the intended way.

On system testing, a testing dataset comprising of selected students was selected and later fed to the trained model. The selected classification algorithm is supposed to classify the data into 3 major categories which are good, Average and poor.

### 4.2.1 The Hardware Used for Testing

The researcher used an HP laptop with the following specifications

* Intel’s 6th generation core i7 @3.2 GHZ
* 8GB of physical memory  Windows 10 pro 64-bit

### 4.2.2 Conditions for testing

The obtained testing results were based on the following conditions:

* Dataset is divided into 2 with 70% used for model training and 30% testing purposes.
* Training of the classifier should be done in the same environment for efficiency.
* The Algorithm’s confidence level being set to less than 95 in all test cases

4.3 Presentation of results

### 4.3.1 a summary table of selected attributes of dataset

|  |  |
| --- | --- |
| **Column name** | **Description** |
| Student\_ID | Student Registration Number |
| Student Name | Name of Student |
| Gender | Male/Female |
| Age | Date of birth |
| Tuition Fees Statement Average | An average of the rate at which tuition  fee has been paid over a period of time |
| Dining hall food purchase statement average | An Averaged value of food purchase statement |
| Enrolment year | First year at university |
| Parent/guardian | Both parents, single parent or orphan |
| Status | Married or single |
| Place\_Of\_Birth | Place of birth |
| Student department | Student branch (e.g. Sciences) |
| City | Location |
| job of father | job of father |
| Student account Summary | A brief statement about the student by the student |

Figure 14 summary table of selected attributes of dataset

I removed some irrelevant attributes, and then processed the selected attributes via rapid miner IDE. Only the attributes that have importance in prediction of students' financial backgrounds are selected. These attributes are related to e.g. the job of father, parent/guardian, Tuition Fees Statement Average, Dining hall food purchase statement average and so on. Data reduction was done by selecting the most important attributes without losing quality.

To apply the classification algorithm on rapid miner IDE, the data set had to be split into training and testing data. In this study, the data set was divided into 70% training data and 30% testing data, then, the labelled features have been selected to apply KNN. Precision, recall and accuracy of results are measured for the applied algorithm. The accuracy, recall, and precision are used as measurements for the performance of the classifier. Precision means the proportion of data which is classified correctly. And recall means the percentage of information relevant to the class and its correct classification. While accuracy is a percentage of instances classified correctly by classifier. A high percentage reflects a strong relation between the features used in training process.

### 4.3.2 Comparing Error Rate with the K Value

the performance of the naive KNN is quite sensitive to k, the size of the neighbourhood, and it generally favours small k. Using all data sources, naive KNN has prediction accuracy of 75%, 80.5%, 76.2% and 89.9% for k = 5,8, 10,18 respectively.

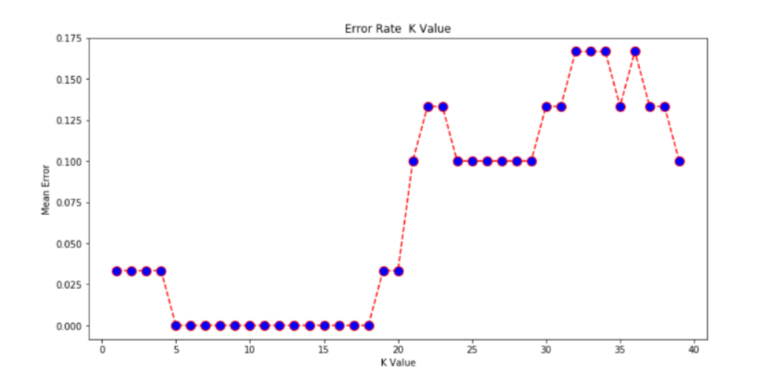


Figure value range

From the output it can be see that the mean error is zero when the value of the K is between 5 and 18. Playing around with the value of K impacts the accuracy of the predictions.

### Performance Testing

Performance testing is in general, a testing practice performed to determine how a system performs in terms of responsiveness and stability under a particular dataset. It can also serve to investigate, measure, validate or verify other quality attributes of the system, such as reliability and resource usage. The performance of the system was evaluated against two (2) major errors that a system may exhibit:

#### False acceptance rate

This is the rate at which the system identifies or categorize a data attribute into a not suitable category or class.

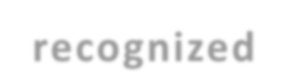
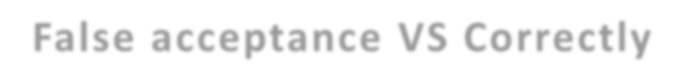
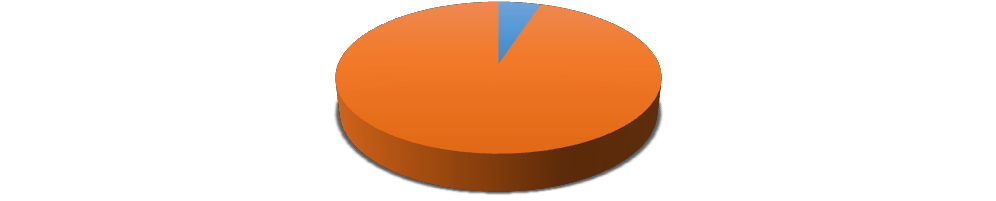
In this system it means that a person’s data will have been placed into the wrong category after processing. However, it is important to bear in mind that this error must be kept at very minimum because if the rate becomes high it means that the system is not fitting its purpose which is a type 1 error (False positive). False acceptance rate is calculated using the following formula:

FAR = Number of false accepted faces / Number of testing datasets

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Number**  **datasets** | **of** | **Testing** | **Number of training**  **datasets** | **False accepted** | **FAR** |
|  | 20 |  | 100 | 1 | 5% |

Figure False acceptance rate

The above table shows the false acceptance rate from the system. Using the obtained results, the system throws a high false acceptance rate when we have a small training dataset and on the other hand produces a lower false acceptance rate when a larger training dataset is used.



**False**

**acceptance**

**VS**

**Correctly**

**classified**

5

%

95

%

Figure False Acceptance Vs Correctly classified

The table shows the false acceptance rate of the system as a percentage. The system attained 95% effectiveness in classification and the false acceptance rate was 5% which is very small for the overall system. Quantitative statistics obtained on False Acceptance Rate shows that the system is not much prone to false Acceptance Rate since the rate is low.

#### False Rejection Rate

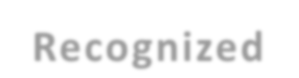
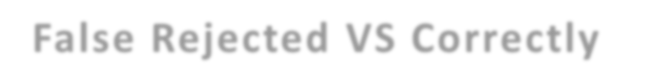
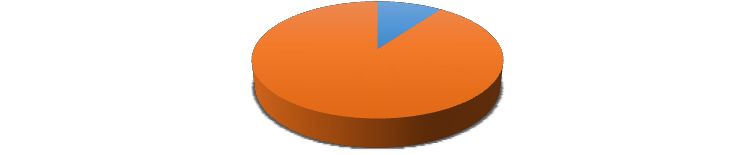
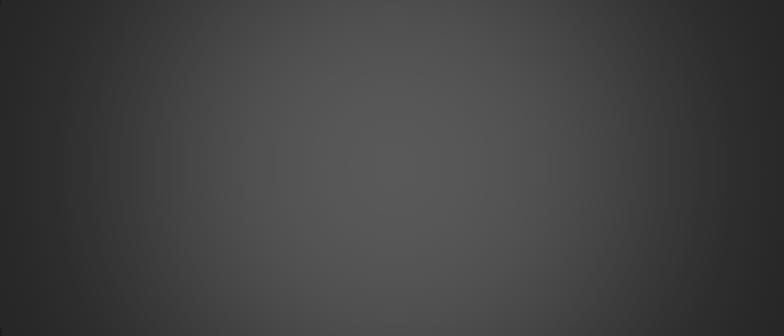
This is the rate at which a data attribute is rejected and is regarded to be not part of a certain category which it belongs to. This is also known as Type II error or False negative calculated as:

FRR = Number of False Rejected / Number of testing datasets

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Number**  datasets | **of** | **testing** | **Number of training**  datasets | **False Rejected faces** | **FRR** |
| 20 |  |  | 100 | 2 | 10% |

Figure False Rejection Rate

The above table shows the false rejection rate from the system. Using the obtained results, the system throws a high false rejection rate when we have a small training dataset and on the other hand produces a lower false rejection rate when a larger training dataset is used.



**False**

**Rejected**

**VS**

**Correctly**

**Classified**

10

%

90

%

Figure False Rejection vs Correctly Classified

The system attained a false rejection of 10% which means only 2 datasets were rejected and 18 were correctly categorized which represents the largest portion of the population sample.

#### Computational time

CPU time (process time) is the amount of time for which the central processing unit was used for processing instructions of a computer program or operating system as opposed to for example. waiting for input/output operations or entering low-power mode. CPU time is measured in seconds calculated as follows:



Computational time obtained means that the greater number of training data available in the database, the more computational time needed but generally the system is efficient.

|  |  |  |
| --- | --- | --- |
| **PC used and Specifications** | **Time for training dataset model** | **Computational Time For**  **Classification** |
| Hp 5th gen laptop @3.1GHZ | 5.0 seconds | 4.9 seconds |
| Hp 6th gen i7 Desktop @2.7 | 2.0 seconds | 1.3 seconds |

Figure Computational Time

The table from the above shows the computational time for the system in training the dataset model and classification given the certain conditions for testing stated earlier on. According to the above results in the table, more CPU cores will usually lower the computational time for both training and classification as compared to the CPU with fewer CPU cores.

* 1. Summary of results

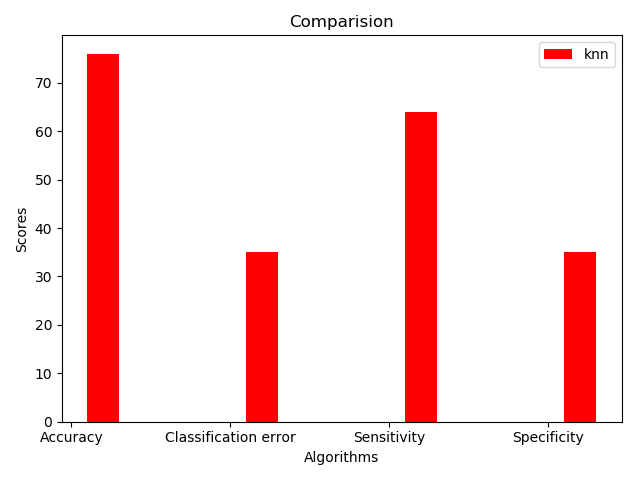


Figure 21 KNN classification Results

According to the obtained results system is efficient because the system is able to classify student profiles according to their financial backgrounds with a success rate of 89%. The prototype has yielded good results in terms of computational time of training the model.

4.5 Conclusion

Basing on the results obtained in this chapter the introduction and use of tuition fees crowd funding system can go a long way in assisting students to attain financial assistance more efficiently.Educational databases contain large amounts of data that is increasing rapidly. To understand more about student data, classification algorithms can be applied to the educational datasets. The study focused on KNN as a major classification algorithm to come up with a student financial background prediction model. In our presented study, KNN algorithm had an accuracy of 89% in classification which means a strong relationship between the student data and their backgrounds. As future work, more classification algorithms can be applied on different educational datasets to come up with predictions that can help in decision making. The next chapter contains major conclusions drawn by the researcher and future recommendations.

# CHAPTER FIVE: CONCLUSION AND RECOMMENDATION

5.1 Introduction

This is the last chapter consisting of the conclusions and the recommendations that can be enhanced on the tuition fees crowd funding system for it to completely fit its purpose. In this chapter the research objectives are going to be used to conclude and review the importance of this research by simply comparing the system objectives against the outcome of the results produced by this system. This also helps to check if the research objectives are met or not.

5.2 The reviewing of the Aims and Objectives

The main aim here is to design and implement a tuition fees crowd funding system which will help in assisting tertiary students financially. After the implementation, the prototype has been tested in different scenarios and has yielded a favourable result of 89% in terms of accuracy in data classification. The accuracy of the system was based on the data classification rate and accuracy as well as the efficiency which was determined by the computational time of the system in performing certain tasks. A tuition fees crowd funding system prototype which uses artificial intelligence to classify students according to their financial backgrounds was developed. The system is able to perform its processes while keeping the computational time and the false acceptance error rate at low percentage.

The researcher developed the system using latest computing resources (hardware) and also viable software resources such as the python sklearn, numpy and pandas modules which comprises of efficient algorithms for data classification which resulted in the efficiency of the system.

5.3 Challenges faced

Many researchers have encountered problems that are prone to data classification systems, in the implementation of this system the researcher also encountered some challenges. The system was affected by the quality of the data obtained and for this matter the classification results were affected in some cases.

5.4 Future Work

This system was implemented using python module which are sklearn, pandas and numpy which offer frameworks for data classification taking advantage of the CPU of the machine only. However, such system can be implemented using some other algorithms such as the naïve bayes classification algorithm or the neural networks which take advantage of both the CPU and the GPU of the PC.

5.5 Recommendations

The implemented system can classify vast amounts of student data into categories which allow us to determine their financial backgrounds. The system was implemented taking advantage of the capabilities of the KNN algorithm. The classification accuracy obtained after the tests shows that the algorithm can be used in other several real time environments which work with data classification.

# References

[1]. D.T. Larose. “Discovering Knowledge in Data, An Introduction to Data Mining”. Wiley Interscience, pp. 90-106. 2005

[2]. R. Agrawal. “K-Nearest Neighbour for Uncertain Data”. International Journal of Computer Applications (0975-8887). 2014. Vol. 105 No. 11 pp 13-16.

[3]. Parvin, Hamid, Alizadeth, Hoseinali, and M. Behrouz, “A Modification on K-Nearest Neighbour Classifier”. Global Journal of Computer Science and Technology. 2010. Vol. 10 No. 14, pp. 37-41.

[4]. Patil, R. Tina, and S. Sherekar, “Performance Analysis of Naïve Bayes and J48 Classification Algorithm for Data Classification”. International Journal of Computer Science and Applications. 2012. Vol. 6 No 2, pp. 256-261

[5]. Zha, Qiang, and D. Wang. "Case Study: The Chinese Government Scholarship Program—the Brain Development Scheme That Illuminates a Vision Across 30 Years." (2018).

[6]. [Dominique, G.,Huck, N. (2005) On the Use of Nearest Neighbours in ﬁnance. http://hal.archives-ouvertes.fr/docs/00/18/08/58/PDF/GueganHuck\_ﬁnance2005.pdf [Accessed: 21st November 2019].](http://hal.archives-ouvertes.fr/docs/00/18/08/58/PDF/GueganHuck_ﬁnance2005.pdf%20%5bAccessed:%2021st%20February%202014%5d.)

[7]. [15] D. Kanojia, “Comparison of Naive Bayes and K-NN Classifier,” vol. 65, no. 23, pp. 40–45, 2013.

[8]. Dudani, S. A. (1976) „The distance-weighted k-nearest neighbour rule‟, IEEE Transactions on Systems, Man and Cybernetics, vol. 6, no 4, pp. 325–327.

[9]. Ishii, N., Hoki, Y., Okada, Y. and Bao, Y. (2009) „Nearest neighbour classification by relearning‟. In: Proceedings of the 10th international conference on Intelligent data engineering and automated learning (IDEAL'09), pp., 42-49.

[10]. Kotsiantis, S., Pierrakeas, C. and Pintelas, P. (2003) „Preventing student dropout in distance learning systems using machine learning techniques‟, In: Proceedings of Seventh International Conference on Knowledge-Based Intelligent Information & Engineering Systems, Lecture Notes in Artificial Intelligence, Vol. 2774, Springer-Verlag, pp. 267–274.

[11]. Islam, M. J., Wu, Q. M. J., Ahmadi, M. and Sid-Ahmed, M. A. (2007) „Investigating the Performance of Naive-Bayes Classifiers and K-Nearest Neighbour Classifiers‟, International Conference on Convergence Information Technology, pp.1541–1546.

[12]. Cover, T. and Hart, P. (1967) „Nearest neighbour pattern classification‟, IEEE Transactions on Information Theory, vol. 13, issue 1, pp 21–27.

[13]. Altman, N. (1992) „Introduction to kernel and nearest-neighbour nonparametric regression‟, The American Statistician, vol. 46 no. 3, pp. 175–184

[14]. Cover, Thomas M.; Hart, Peter E. (1967). "Nearest neighbour pattern classification" (http://ssg.mit.ed u/cal/abs/2000\_spring/np\_dens/classification/cover67.pdf) (PDF). IEEE Transactions on Information Theory. 13 (1): 21–27. CiteSeerX 10.1.1.68.2616 (https://citeseerx.ist.psu.edu/viewdoc/summary?doi =10.1.1.68.2616). doi:10.1109/TIT.1967.1053964 (<https://doi.org/10.1109%2FTIT.1967.1053964>).

[15]. Agrawal, A., C. Catalini, and A. Goldfarb. 2011. Offline Relationships, Distance, and the Internet: The Geography of Crowdfunding. Cambridge, MA: National Bureau of Economic Research.

[16]. Vance, S., (2013). Crowdfunding Success – A closer Look at Economic Models. Spinnakr Blog. http://spinnakr.com/blog/ideas/2013/07/crowdfunding-success-a-closerlook-at-economic-models/

[17]. Brabham, D. D. (2008) Crowdsourcing as a Model for Problem Solving: An Introduction and Cases. Convergence: The International Journal of Research into New Media Technologies. 14(1), pp. 75-90. (2008) London, Los Angeles, New Delhi and Singapore: Sage Publications.

[18]. Mollick, E. (2013) The Dynamics of Crowdfunding: Determinants of Success and Failure. The Warton School of the University of Pennsylvania. Available: ssrn.com/abstract=2088298 Accessed: May 6, 2013.

[19]. Sharma, A. (2010) Crowdsourcing Critical Success Factor Model - Strategies to harness the collective intelligence of the crowd. Working Paper 1 – 2010. Available at: http://irevolution.files.wordpress.com/2010/05/working-paper1.pdf Accessed: May 7, 2013

[20]. Westerlund, M.; Singh, I.; Rajahonka, M.; Leminen, S. Can short-text project summaries predict funding success on crowdfunding platforms? In ISPIM Conference Proceedings; The International Society for Professional Innovation Management (ISPIM): Manchester, UK, 2019; pp. 1–15.

[21] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. Journal of Machine Learning Research, 2003.

[21] E. Gilbert. The language that gets people to give: Phrases that predict success on Kickstarter. In Proceedings of CSCW,2014

[22] Alegre I, Moleskis M (2019) Beyond financial motivations in crowdfunding: a systematic literature review of donations and rewards. Voluntas. <https://doi.org/10.1007/s11266-019-00173-w>