Learning the Student’s Happiness Model
Project Progress Report

Submitted by: Nwamu Phillippa Unoma
Supervisor: Dr. Xiaojun Zeng

School of Computer Science
The University of Manchester

5/10/2013
## Contents

Abstract ............................................................................................................................................... 2  

Chapter 1: Introduction ...................................................................................................................... 3  
1.1 Project Aims and Objectives ....................................................................................................... 4  
1.2 Project Scope ............................................................................................................................ 5  
1.3 Report Structure ......................................................................................................................... 5  

Chapter 2: Background and Literature Review ................................................................................... 6  
2.1 Overview ..................................................................................................................................... 6  
2.2 Happiness Economics .................................................................................................................. 7  
2.3 Subjective Well-being and Happiness ....................................................................................... 9  
2.4 Concept and Definition of Happiness ....................................................................................... 10  
2.4.1 Pursuit of Happiness ........................................................................................................... 11  
2.4.2 Causes of Happiness .......................................................................................................... 12  
2.4.3 Measures of Happiness ....................................................................................................... 13  
2.5 Happiness in Schools ............................................................................................................... 14  
2.6 Related Works ......................................................................................................................... 15  
2.6.1 Review of Related Works .................................................................................................. 15  
2.6.2 Review of Technical Approach of Related Work ............................................................. 17  
2.6.3 Weakness and Limitations of Reviewed Works ................................................................. 19  
2.7 Background Summary ............................................................................................................. 20  

Chapter 3: Project Progress .............................................................................................................. 21  
3.1 Research Methodology ............................................................................................................. 21  
3.2 Evaluation Plan ......................................................................................................................... 22  
3.3 Collaborative and Communication tools .................................................................................... 22  
3.4 Project Plan ............................................................................................................................... 23  
3.4.1 List of Project Milestones .................................................................................................. 23  
3.4.2 List of Project Deliverables ............................................................................................... 23  
3.5 Progress Report ......................................................................................................................... 24  
3.6 Project Progress Summary ....................................................................................................... 31  

References .......................................................................................................................................... 32  

APPENDICES ...................................................................................................................................... 38
Abstract

Happiness is an emotional feeling experienced by individuals. It has a positive impact on an individual resulting in an increased mental strength to handle problems, greater self-confidence, sound mind and improved performance and well-being. Little research has been carried out on the study of student’s happiness, hence the motivation for this study. The purpose of this project is to study the factors which influence the level of happiness of students in order to produce a mathematical relationship to model the happiness of students. This study employed survey method using quantitative approach to gather and analyse data. From various interviews with students and review of previous studies, the factors which influence the level of happiness of students were identified. Based on these factors a questionnaire was then created and distributed to students. Data collected formed the dataset which was used for analysis. Data analysis would be done in two stages; the first stage is an overall analysis of the identified factor. The second stage is the segregation stage which involves segregation of data into clusters and analysis being done on the identified clusters. In order to develop the mathematical model of student’s happiness, different machine learning techniques would be applied to the dataset. The accuracy of each machine learning technique would be evaluated and the best method would be chosen to develop the model. The results gotten from this project would help educational institutions and governments in enhancing the quality of education by finding ways to increase the happiness of students in educational institutions. This would result in better performance of students as well as increase the level of interest of students in Education.
Chapter 1: Introduction
Since the beginning of civilization, the value of human life has been discussed by many scholars. There have been different opinions by individuals of an ideal state and some are of the opinion that having a lot of money constitutes an ideal state, to others experiencing meaningful relationships. These individuals might share a subjective feeling of wellbeing yet they differ in external conditions (Eddington and Shuman, 2005). There have been an increasing number of interests to the scientific study of subjective wellbeing this is shown by the growing number of papers on subjective wellbeing (Kahneman and Krueger, 2006). A lot of theories about subjective wellbeing have been proposed and subsequently tested, with the most theoretical advancement shifting importance from identifying the needs which needs to be satisfied in order for an individual to be happy to identifying factors which would influence the happiness of an individual (Diener et al., 1999). Life satisfaction and happiness are important goals for most people hence subjective wellbeing researchers place importance on recognising the processes which causes happiness.

There have been various concepts of happiness which include positive emotions, a feeling of contentment, life satisfaction, pleasure, a meaningful life etc. Happiness does not necessarily mean being overcome with joy every time of the day, but frequently experiencing more positive emotions compared to negative emotion and it is used in assessing the satisfaction or sense of well-being with regards to an individual’s overall life. The name ‘Subjective well-being’ (SWB) is the psychological name for ‘happiness’ which would be used interchangeably throughout this study and it represents an individual’s assessment of their lives – consisting of affective assessments (emotions and mood) such as negative and positive psychological emotions; and cognitive feelings such as life satisfaction. If individuals are contented with their life situations they are said to have great SWB and have regular positive feelings and less negative feelings (Eddington and Shuman, 2005). In a study by Ahn et al, (2004) happiness was said to be more important than money with most individuals highlighting its significance as a key component to physical, emotional and psychological wellbeing. As a result, lots of people spend most of their time in the pursuit of happiness (Argyle, 2001).

Rapid universal growth has been made possible by recent advancements in technology especially in the information and communication technology. These advancements have made it possible for the educational system to benefit in preparing students for the future as such the quality of the educational system should be dynamic, flexible, versatile and yet ready to progress towards excellence. One of the tasks of the educational system is making learning experience exciting for students. In the work by (Goodlad, 1984) he puts forward the question ‘why our school is not a happy place’. He doubts that students would have any indication of happiness in their list if asked to list the qualities of school. If students are expected to enjoy the learning process there have to be a significant re-engineering of the curriculum and learning instructions. Schools are places were individuals are educated and not only a
place for staff examiners and employment. To help promote student’s happiness, in-school happiness can be explained in terms of positive feeling like enjoyment and pleasure.

Noticeable reforms have been made by governments and university authorities over time to improve the quality of education. However, with the growing importance on measuring student performance primarily through their academic results, little effort and deliberation have been given to student’s happiness either by trying to satisfy or increase it. The importance of happiness is visible in different domains of life. It is found that individuals who are happier tend to live longer, have more fulfilling interpersonal relationships with others, earn more money, and are more effective at work (Lyubormirsky et al, 2005). The significance of happiness in school as well as in the community and home was shown in a research conducted with adolescence (Huebner et al, 2006) and this research showed that lack of happiness is a vital risk factor for detachment from school, depression, as well as loss of social support from parents and friends. As Noddings (2003) suggests in her research, the key goal of education should be student’s happiness, since happier students learn better than unhappy student.

There are lots of factors which contribute to an individual’s happiness, this study analyses and identifies the factors influencing the happiness of student’s in the universities and determining the extent to which each factor plays a role in the happiness of a student. Earlier research on student’s happiness identifies factors such as educational performance and environment, resources, social relationships, and extracurricular activities. This study provides more insight to happiness of university students by identifying the factor that influences student’s happiness across different nationalities, genders, age and faculty. This study takes a quantitative approach; data is gathered from students in various universities from surveys sent out and the data is analysed using computational and machine learning methods which would build a mathematical model which analyses and identifies the quantitative impact and strength of each factor on student’s happiness.

Happiness usually has a positive impact on an individual resulting in improved self-confidence, enhanced mental strength to handle difficulties, improved wellbeing and performance. Therefore the results gotten from this study would serve as a way to help government and academic institutions not only in their bid to enhance the quality of education but also aid in understanding and improving the happiness of students.

1.1 Project Aims and Objectives
This project aims to identify the factor which influences student’s happiness and to build a mathematical model of student’s happiness. By employing machine learning methods and computational intelligence on the data collected. In order to show the relationship between the factors and the level of happiness of students and also the extent to which each factor plays a role in student’s happiness.
In order to achieve the aim of the project the following objectives have been set out.

- To analyse and identify the factors which impacts on the student’s happiness.
- To form the dataset which would be used for learning by designing and collecting at least 300 questionnaires from various students.
- To determine the appropriate methodology to be used for learning
- To learn from the dataset by applying computational intelligence and machine learning methods to identify the student’s happiness model.
- To identify the top factor impacting the happiness level
- To identify the accuracy of the learning techniques employed
- To analyse the result and give appropriate recommendation

1.2 Project Scope
In order to achieve the project’s objectives and prevent unbounded research the scope of this project has to be made clear as follows

- Data is gotten from only current university students in any part of the world
- Both supervised and unsupervised machine learning methods are used
- Only machine learning methods capable of dealing with numerical attributes are used

1.3 Report Structure
This report started with the introduction as well as the project aims, scope and limitation, the rest of the report is structured as follow:

Chapter 2: Background and Literature review – This chapter covers previous research works on student’s happiness and provides a wider view of happiness economics as well as an analysis of the approach used by previous researchers.

Chapter 3: Project Progress – This chapter shows the research methodology to be employed as well as the evaluation plan, project plan and the progress made on the project so far.
Chapter 2: Background and Literature Review

2.1 Overview

There have been lot of studies on subjective wellbeing and its components such as life satisfaction and happiness over the last four decades. Numerous contributions have been made by Ed Diener in this field, his works being cited in hundreds of papers. There has been progress in understanding what makes up subjective well-being (Diener, 1984) with increasing importance being placed to subjective well-being by Individuals. The study of wellbeing is being classified under behavioural science as it studies individual’s evaluation of their life (Diener et al., 2006).

Despite hardships and problem subjective wellbeing and happiness are reported from about eighty to eighty five per cent of the world’s population (Scollon et al., 2003; Seligman et al., 2005). Research has shown that past, present and future lives of an individual tend to be thought about positively than their lives are in reality. This makes for a positive thinking and better adaptive day to day behaviours (Fredrickson and Joiner, 2002). Happiness is encouraged by positive thinking and adaptive behaviour and is a characteristic for more successful and beneficial feelings which enhances better coping (Lyubormirsky et al., 2005).

The psychological term for subjective well-being is happiness (Eddington and Shuman, 2005); generally both terms have been used interchangeably. Due to the growing recognition of its positive benefits subjective well-being has become important to stakeholders such as Individuals, government, philosophers, academics, policy makers, economists, organizations and nations. There have been several uses of data collected on subjective wellbeing emphasizing its increasing importance. Nations are not only being ranked by economic indicators such as gross domestic products, housing starts, inflation etc. but also by the level of happiness (Diener and Suh, 1997). Economists have made lot of use on subjective wellbeing data, exploring both micro and macro oriented questions. Using this data the relationship between happiness and economic growth was examined by Easterlin (1995) also data on subjective well-being was used to offer a peripheral check for economic indicators (Nordhaus, 1998; Krueger and Siskind, 1998) amongst others. As many Individuals have discovered; living lavishly or earning wonderful salaries does not guarantee happiness. It is therefore important that public policies put more emphasis on well-being as a means of increasing happiness level of Individuals, than putting all their efforts on improving the national income. Questions have been raised over the amount spent improving national income and strengthening the economy which cannot be justified when compared to the level of happiness of Individuals (Layard, 2003). Inglehart and Klingemann (2000) suggested that as individuals basic needs are met; there is a shift to post materialistic stage in which they are involved with self-satisfaction.

The economics of happiness does not claim to substitute income-based measures of well-being rather it aims to complement them with measures of a broader range of well-being. These measures are
based on surveys conducted on a large scale from a large number of Individuals across nation over time. These Individuals are asked to evaluate their own well-being. The results gotten from these surveys offer information on the significance of a range of factors which impact well-being like earnings as well as others such as civic trust, health, employment position and marital status. Hence the motivation behind happiness research is not to replace current measures of assessing the life satisfaction or well-being of Individuals, rather provide a complementing technique for more rounded results.

2.2 Happiness Economics

Until recently, very little empirical research has been conducted on the subject of happiness in economics. In general, the study of economic psychology experienced a fallow period in the mid-20th century primarily because the economic methods and macroeconomic models of that era were rigid and inflexible. Economists faced challenges when evaluating the efficacy of public policies or economic interventions on the polity and often had to default to established economic measures to measure public well-being. Within the scientific society, the importance of the study of how happiness changes with social or economic factors has been slowly gaining recognition since the 1980s. Within the last decade in particular, there has been an explosion in the adoption of the self-reported happiness scale. That said Easterlin (1974) is widely credited as being the pioneer of the field of happiness economics. He conducted extensive research to determine the correlation between income levels of individuals and their self-reported happiness. He found that intra-country comparisons display a strong mostly positive relationship while international comparisons indicate no clearly obvious relationship. Easterlin (1995) strongly argued against the concept that the ‘keys to happiness’ lies in an increase in income, arguing instead that a rise in one’s income relative to their counterparts is more important. There exists an income/happiness plateau, Ng (1997) argues that once the point at which one’s income levels adequately take care of their basic needs has been attained, additional income does not necessarily lead to additional happiness. In fact, he argues further that diminishing returns set in after that point has been attained and further increasing an individual’s consumption power could actually lead to reduced happiness by predisposing them to weight-related illnesses, increased stress levels etc. This relative income concept forms the basis of the ‘Easterlin Paradox’ (Easterlin, 1995), which is the core of happiness economics research.

Employing time series data from post reunification East Germany, Frijters et al (2004) tested this concept and discovered that particularly in the immediate post-reunification years; there exists a direct relationship between levels of income and happiness such that an increase in the former would bring about an accompanying increase in the latter. Clark et al (2008) argue that, when compared to particular points within a time period, aggregate time series slightly masks the actual relationship between happiness and income. When analysed, they report an inverse relationship between income and happiness in that an increase in the former brings about a decrease in the latter, thus in the
income-happiness space, the gradient of the curve is reduced. While in actuality, at any point in time the richest individuals display above average happiness while the poorest report less than average happiness. Unfortunately, the time series aggregate data inadvertently inverts this relationship and significantly undermines its reliability when explaining happiness economics. Di Tella and MacCulloch (2008) present an alternative ideology that states that the perceived insufficiency of an individual’s absolute income in determining their level of happiness is not paradoxical but rather an omission of more significant microeconomic indices.

There is a perpetual trade-off between the amount of time spent engaged in productive labour and engaging in satisfaction, leisurely activities. The premise is that a higher income encourages individuals to expend more time working, which reports a positive slope on the labour-supply graph. However, individual preference is an important input for the work-leisure trade off, studies suggest that income is not necessarily as critical a determinant in the equation and its use as the sole economic indicator is an insufficient measure of an individual’s happiness. Several studies have evaluated the importance of relative/absolute consumption (Alpizar et al., 2005), for instance the relative income hypothesis states that an individual’s approach to saving and consumption is less influenced by absolute level of consumption and more by their income in relation to others. This is the basis for the so-called keeping up with the Joneses mind set.

Headey et al (2008) analysed data representing subjective well-being from individuals in five different countries in order to determine whether, when compared to income as a standalone indicator, wealth and consumption contribute significantly more to power in lifestyle satisfaction. Unfortunately, the consumption data is only available for two out of the five countries, Britain and Hungary, so the results should be analysed in that context. The result findings suggest that examining an individual’s income as a single determinant of happiness could be misleading and when wealth and consumption are included, income constitutes just a fraction of the life satisfaction variance. An OLS regression assessment of the objective controls reported wealth to be somewhat at par with income, while the consumption depends on the country being analysed and seeing as consumption data is not available, it is practically impossible to arrive at a firm conclusion.

The findings on happiness have also been applied to the field of unemployment studies to evaluate the pressure exogenous economic or financial factors place on the subjective well-being of individuals without paid employment. Analysts have attempted to measure whether the idleness that accompanies unemployment causes unhappiness or whether individuals simply choose unemployment in order to benefit from government’s welfare programs (if any). Clark and Oswald (1994) examined the relationship between unemployment and happiness and found the former to be statistically significant on the one hand and negatively correlating with the latter on the other. Furthermore, factors such as the individual’s age, background and highest attained level of education could introduce variance to
the findings, so these factors were disaggregated in order to report what percentage they have and what roles they play in lowering the subjective well-being of the individual. It was discovered that individuals with a higher level of education and in their mid-30’s without paid employment were the least happy. Di Tella et al (2001) corroborate these findings, they suggest that lower levels of unemployment and an accompanying low level of inflation has a generally positive effect on individual’s happiness. Blanchflower and Oswald (2004) conducted a comprehensive study that further validates these findings and even goes a step further by suggesting monetary sums for different levels of happiness as means of compensation. They estimate that, amongst American men, a yearly rise in income of about $60,000 is required to fully compensate.

2.3 Subjective Well-being and Happiness
Subjective well-being is made up of an Individual’s affective and cognitive assessment of his or her life as a whole (Snyder and López, 2009). These assessments consist of cognitive judgements of fulfilment and satisfaction. Scholars of subjective well-being presume that one of the factors of a good life is if an Individual likes his or her life. Subjective well-being is a wide concept which includes undergoing great level of life satisfaction, small levels of negative moods and emotions and great levels of pleasant moods and happiness.

Different characteristics such as nonattachment, wisdom and love have been suggested by religious leaders and philosophers as being the fundamental elements of a contented existence (McMahon, 2006). However utilitarian argue that the absence of pain and the presence of pleasure are the crucial qualities of a good life. Hence the intellectual fore bearers of subjective well-being researchers were the utilitarian, concentrating on the emotional, mental and pain and physical pleasures that is experienced by individuals. According to Ryan and Decci (2001); Ryff & singer (1998), one main factor of a good life possessed by an individual is abundant joy, though there exists further desirable personal factors beyond whether an individual is happy.

Empirical studies of subjective well-being started to take form in the early twentieth century. Flugel studied Individual’s mood as early as 1925, by having individuals record their emotional experiences and summing up their emotional responses across moments. His works were the forerunner of current sampling methods to assessing subjective well-being online as individuals go about their normal lives (Snyder and López, 2009).

There are two methods of assessing the well-being of an Individual, either objectively or subjectively. In subjective method the well-being of an Individual is assessed based on an Individual’s personal values and preferences such as likes and dislikes or aspirations and desires. However objective theorists disagree and reject this method arguing that there are other things in life which may not be valued by an individual but are objectively significant. These two methods can be used to explain further the views on happiness.
2.4 Concept and Definition of Happiness

Happiness has been defined differently by various people from different locations. It is commonly used by an individual to show a condition of joy with themselves and events surrounding them. Happiness was described by Veenhoven (1984) as the level to which an individual positively assesses the overall quality of his existing life-as-a-whole. In another word, how much an individual prefers the lifestyle he/she leads. Lots of concepts have been put forward with regards to how happiness can be considered. Happiness (eudaimonia) was described by Aristotle as the capability to achieve one’s potential (Waterman, 1990). While happiness was described by Bentham as the absence of pain and presence of pleasure (Tatarkiewicz, 1976). Modern theorists came up with their own concepts explaining happiness as the result of fulfilling certain needs in one's life.

Happiness consists of 'contentment' and 'hedonic level of affect' (Veenhoven, 2010). The level at which an individual is affected by pleasure is described as the hedonic level of affect. The regularity of this level of affect can be identified over a period of time. Emotions of tranquillity or joy are typical example of such affects. Alternatively contentment can be presented as how an individual weigh the success of predetermined objectives over a defined length of time, generally combining the Future and past. This idea is very subjective since individuals are accountable for identifying the levels of their personal accomplishment which they are satisfied with. There are debates which exist over the assigned weights given to each components in assessing life overall, since it combines both concepts. Which concepts bear more value the hedonist level of affect or contentment?

According to (Brülde, 2007) there are four different concepts of the views on happiness:

- **Hedonistic theory**: The hedonist theory view happiness as an excellent measure of pain and emotions of enjoyment.
- **Cognitive view**: The cognitive view of happiness is subjective and is related to the concept of contentment by Veenhoven (1984). It is the constant positive disposition of life by an individual without actually feeling any joyous or happy emotions.
- **Mood Theory**: In this view happiness is characterized as being in a positive or optimistic mood without any occasion actually leading to this mood. Sometimes this view is seen as a limited side of the hedonistic concept.
- **Hybrid Theory**: This theory views happiness as a complex mixture of both cognitive concept and level of affect. It is roughly the cognitive assessment of one’s life and consequently having a pleasant feeling. The level of happiness of an individual comprises of two factors (1) how pleased an individual feels with his life, the positive attitude towards his life and (2) How good an individual feels about this (Brülde, 2007).

From a scientific view some researchers believe that having large quantity of DNA is a determining factor of happiness. According to Lykken and Tellegen (1996) genetics set an established level of
happiness, which means that an individual has a determined highest level of happiness regardless of any pleasant or unpleasant circumstance. This established level can be as high as fifty per cent to eighty per cent and it is passed down along generations (Sheldon, and Lyubomirsky, 2006). This view is said to be false, as the level of happiness of different individuals varies and is never constant due to the ever changing circumstances. However there is a general consensus that environmental factors have an impact on the level of happiness of an individual. An Individual’s level of happiness tends to be affected by the environment in which they come from or live in.

2.4.1 Pursuit of Happiness

In trying to answer the question if people can be happy, it is important we establish the difference between ideal and actual happiness at the onset (Tatarkiewicz, 1976). Happiness that is lasting, complete and touch all aspects of life can be defined as ideal happiness. Such a state of happiness is eternal, perfect, and pure and has incredibly high expectations but it is unattainable by any person. However, individuals might still be able to experience mostly positive emotions and also be contented with their life. Psychologists are interested in actual happiness as a subject of scientific inquiry (Kesebir and Diener, 2008).

In order to shed more light on the possibility of attaining happiness; scientific psychology addresses two relevant questions: Is happiness being reported by Individuals and is happiness a feasible, adaptive and evolutionary occurrence? Globally evidence gathered indicates that the answer to the first question is positive. The evidence available was analysed by Diener and Diener, (1996) concluding that the large majority of people fall in the positive range on the scale of happiness, this includes individuals with obvious drawbacks, such as quadriplegics or those in the smallest income categories. This is corroborated by an opinion poll (Pew Research Middle, 2006) which revealed that eighty four per cent of Americans view themselves as being “very happy” or “pretty happy” (Kesebir and Diener, 2008). Furthermore, eighty six per cent of the forty three countries involved in Diener and Diener, (1996) research had an average level of happiness beyond the midpoint on the happiness scale. Though it is unusual for people to be regularly happy or delighted, most individuals report being happy frequently. This evidence conflicts with the perspective of life as a “vale of tears” and of the society as a “sink of disappointment.” People appear to have a temperament for mild happiness level, which leads us to our second question: What are the adaptive happiness functions? It has been long identified that adverse feelings (e.g., anxiety, rage, and worry) enables an individual to concentrate on the immediate risk or problem, thereby causing evolutionary fitness. However recently the adaptive benefits brought about by positive feelings are being understood. Fredrickson (1998) suggested that positive feelings enable people to extend their repertoires of thought– action and develop perceptive, physical, intellectual, social, and psychological resources eventually. It follows that happiness is not just a phenomenon, from an evolutionary perspective it is also adaptive and results in various benefit.
A psychological research, performed by Loewenstein, Schooler and Ariely (2003), indicates that the continuous evaluation of an individual’s happiness and the conscious pursuit of happiness may actually be unhealthy to an individual’s well-being. The finding from this research showed that happiness is associated with low levels of self-focused interest (Green et al., 2003). Yet simultaneously the effectiveness of ways to raise happiness is supported (Fordyce, 1977; Lyubomirsky et al., 2005). This implies that whereas being obsessive and self-conscious about happiness may backfire, there are still certain actions people can knowingly undertake or changes in lifestyle that they can purposely create that will raise their level of happiness, such as counting one’s blessings and meditation.

Evidence gathered shows that, even though adaptation undoubtedly happens to some extent and personal ambitions do grow and adjust, individuals do not completely and/or quickly adapt to everything (Diener et al., 2006). For example in a 15-year longitudinal study (Lucas et al., 2004) it was observed that individuals who had experienced widowhood or unemployment, on average did not recover fully and return to their earlier level of happiness. Other research show that hardly, if ever do individuals adapt to certain components in their life like long commutes, interpersonal conflict or disturbance (Haidt, 2006), whereas long-lasting positive results on an individual’s psychological well-being might be felt by other events like plastic surgery (Rankin et al., 1998).

2.4.2 Causes of Happiness

It has been shown that the causes of happiness are complex. Broad research has been conducted on the link between demographic and other environmental aspects with happiness (Cantril, 1965; Bradburn, 1969; Campbell et al., 1976; Inglehart, 1990; Veenhoven, 1995). It was established that happiness is affected by environmental and demographic factors at different levels but to a lower extent than personality. With such factors as age, it was found that life satisfaction do not decrease with age, it often increases with age (Butt and Beiser, 1987; Inglehart, 1990; Veenhoven, 1984; Diener and Suh, 1998). According to a research conducted (Fujita, Diener and Sandvik, 1991), more cases of depression and negative affect are reported by women than men, yet both sex report approximately similar level of global happiness. An explanation might be that women are more likely to admit to negative emotions while men deny feeling such emotions. Studies have shown that personality have an important impact to an individual’s happiness (DeNeve and Cooper, 1998). Constituents of subjective well-being (affective and cognitive) are said to be consistent over time and circumstances so can be predicted reliably from an individual’s constructs and characteristics. Studies show that physical disability, heredity, unemployment and marriage have a causal impact on happiness level (Diener et al., 2000; Argyle, 2003). Cross-cultural research suggests that diverse factors relate with happiness across different cultures (Diener and Suh, 2000). Different factors lead to happiness in persons with unique goals and values. Hence for different people, different strategies would work in diverse situation. Therefore there are various reasons which might cause happiness in
individuals so situational, cultural, genetic, goals, personality and coping strategies must be incorporated.

2.4.3 Measures of Happiness
Considerable amount of time has been spent by psychologists analysing sources of human satisfaction (Argyle et al., 1989; Fox and Kahneman, 1992; Diener and Suh, 2000). They view subjective well-being or happiness as the extent to which individuals view their life in general, or one aspect of their life as being satisfactory. Psychologists believe that asking individuals how they feel would enable them study subjective well-being (Diener et al., 1999). To address ways of measuring an individual’s level of happiness, surveys can be designed. Two approaches have been adopted in measuring happiness- Single-item and multiple-item scales of measurement.

Single-item scale of measurement is known for its brevity and has been considered to be less reliable by psychologists. A typical and commonly used single-item scale is the satisfaction scale which was used in the Eurobarometer scale. Typical questions asked in the latter scale probe the satisfaction levels of individuals, for instance ‘On the whole are you very satisfied’, ‘fairly satisfied’, ‘not very satisfied’, or ‘not satisfied with the life you lead’

In multiple item scale of measurement, random measurement errors are limited compared with single-item scale of measurement; thus increasing the validity of the former. A typical multi item scale is General Health Questionnaire (GHQ) in the British Household Panel Survey which involves the measurement of individual’s subjective well-being. The responses to the following questions are used in measuring GHQ and optional questions measured on a four point Likert scale. The questions are, have you recently:

‘‘(a) been able to concentrate on whatever you’re doing, (b) felt that you were playing a useful part in things, (c) felt capable of making decisions about things, (d) been able to enjoy your normal day-to-day activities, (e) been able to face up to problems, (f) been feeling reasonably happy, all things considered, for positive affect, (g) lost much sleep over worry, (h) felt constantly under strain, (i) felt you could not overcome your difficulties, (j) been feeling unhappy or depressed, (k) been losing confidence in yourself, (l) been thinking of yourself as a worthless person?’’ (Clark and Oswald, 1994)

GHQ and the optional answers provided on the four point Likert scale are used to assess individual’s psychological state or well-being and may indicate negative and positive state respectively. It has been recommended that well-being should not only be measured via surveys but should include real time measurement (Csikszentmihalyi, 1990). Real time measurement may utilise the Experience Sampling Method (ESM) which involves recording periodically individual’s state of happiness using a handheld computer. Recording encompasses set questions answered severally in the course of the
day based on the individual’s activities. This ultimately reduces biases (including memory bias) contained in normal surveys.

A different method—Day reconstruction method (DRM) can be used an alternative to ESM and involves asking participants to complete a diary of their previous day activities which can then be used to assess their happiness (Kahneman et al., 2004).

2.5 Happiness in Schools
In his book titled: "The eternal happiness", Qodsi (2009) deliberated the effectiveness and role of happiness and enjoyment in life, emotions, in creating peaceful environments, and in improving relationships. Schumaker (2009) argues in his book titled "What is happiness" that voluntary activities are related to stress reduction, a stress occurring due to loneliness. Saber, 2009 established in a research titled "Measures to recognize enjoyable schools, perspectives of female students in junior schools of Tehran" that their priorities included such measures as Good teachers, Celebration preparation, Students' participation in executive affairs of school, and Environmental interaction.

Research by Jafari et al (2009) established that individual qualities, instructional factors, socio-cultural as well as participative management are all important factors and are effective in making schools enjoyable however there was found a significant connection between the perspectives of extracurricular instructors and counsellors. In their study, Mahon et al (2010) found that although there was a positive and meaningful difference between happiness and healthiness among all students in general, there was no difference between male and female students’ views about happiness. However, it was found in a research conducted by Oishi, et al (2010) that one's happiness is initiated by an emphasis on individual qualities.

Kelly et al (2009) through a study on the behaviour healthiness among school age children established that students can achieve their identity as well as self-gloriousness through family's high support, school support and also social support. As a result, they will also develop independence, competency and self-awareness. Baumeister et al., 2008 conducted a research to respond to the question if self-esteem will lead to a better performance, individual success, and happiness with a more healthy life and found that self-esteem had a strong relationship with happiness. Crowther (2008) reported that due to the enjoyable curriculum, schools studied in York Shire were very happy. These schools had offered a lot for learning. Rydeen, 2008 quotes Miller's research to reveal that schools have three components to create happiness: a place for children to meet their friends; a place to acquire new knowledge; and a place to love, grow and plan (Talebzadeh and Samkan, 2011 ).
2.6 Related Works

2.6.1 Review of Related Works

One of the earliest works to determine student’s happiness in universities was carried out by Chan et al. (2005) with the goal of identifying the role of social and economic factors in influencing student’s happiness. The study was performed at the University of Western Australia in June, 2003 and has contributed greatly to students’ happiness.

In this study, a survey method which made use of questionnaires to understand the influence of student’s happiness was carried out. Data was collected from over a thousand students attending an economics lecture who participated voluntarily after the end of a lecture. Demographics covered in the questionnaire include course, gender, age; employment history of students while in university, university life and Plans after University. A Likert scale was used previously in measuring satisfaction levels of students (Bradburn, 1969; Easterlin, 1974) a Likert scale (Likert, 1932; Easterlin, 1974, Di Tella, MacCulloch and Oswald, 2001; Arnold, McCroskey and Prichard, 1967). This scale was based on five possible options with 5 = “strongly agree” and at the other extreme 1 = “strongly disagree”.

The dependent variable which was used to assess the overall satisfaction of students who contributed to the survey was on the basis of the answer to the statement “overall I am happy with my university life”. This response was used to assess the satisfaction of students via a survey. In this survey nearly sixty per cent of the students agreed that they were generally happy with the university life. The reliability of the response was verified by an additional statement “Overall university life has been good for me” which was included at the questionnaire end (Chan et al., 2005). Data analysis indicated that sixty six per cent of the responses where valid with significant error.

Results from the survey showed that satisfaction levels do not correlate with gender and may not be an accurate test for happiness (Frey and Stutzer, 2002). When the data was categorized by age, it was observed that high level of satisfaction was recorded by the oldest participants. This was due to the concern over grades faced by new entrants into the university and also concern over future career prospects encountered by older students. Students in their early twenty have the lowest mean level of satisfaction; this was due to their heavy workload. In addition to age other factors such as time management, relationships and income (Easterlin, 1974) may affect the level of satisfaction of students (Chan et al., 2005).

A similar research was performed by Mangeloja and Hirvonen, (2007) to test the factors which impact student’s happiness. The survey method was used consistent with the report of Chan and co-workers (Chan et al (2005)). The data was collected from two hundred and forty six participants from different faculties of the University of Jyvaskyla, Finland. This study attempted to examine the overall
happiness of students vis-a-vis their academic environment. The response to the studied questions was affirmative and the analysis of data is presented.

The comparison of results by Chan and co-workers (Chan et al., 2005) with Mangeloja and Hirvonen, (Mangeloja and Hirvonen, 2007) indicated that majority of students where happy with university life, with Finnish students being considerably happier than Australian students. This finding was confirmed by a reliability question. In terms of age, results obtained showed that gender does not correlate with Finnish student’s happiness similar to the finding of Chan and co-workers (Chan et al., 2005). Factors identified as responsible for the happiness of Finnish students include relationships, personal standards, satisfaction with resources and academic environment. However the students inability to achieve work balance and complete course milestones in due time was shown to affect their academic and general satisfaction levels.

At the University of Algarve in a Marketing class being held at the faculty of economics, a questionnaire was sent out, to get the opinion of students regarding their happiness level. Data analysis revealed that nearly sixty per cent of participants were females and fifty nine per cent were aged between twenty one and twenty five. The level of happiness was rated by asking students to choose on a scale of one to ten how happy they felt based on different activities given. Activities such as going for parties and hanging out with friends were rated high amongst all students. It was observed that females place a higher importance to parties while the males gave preference to activities such as soccer and surfing the net. Data segregation by sex showed that things such as having children, getting married and family was valued amongst the females. The male were more concerned with having a sexually active life (Guerreiro and Vieira, 2007).

Overall most students reported a high level of happiness, though a scale of 10 was not chosen. This goes to substantiate the claim that ideal happiness is unattainable. Further analysis carried out showed that some students where confident and ambitious, while majority proved to be optimistic, tolerant and equally confident. Research results showed that overall most students were happiest when engaging in social activities.

At the Management and Science University in Malaysia, researchers carried out a qualitative study on thirty three students in the medical department. The students were split into 5 groups and were asked questions to get their views on happiness (Al-Naggar et al., 2010). The views on happiness differed across most of the participants. When asked what their main source of happiness was majority of the participants listed money as their top factor. Others factors such as relationships, stable life and good health were also listed. Participants were asked if they would take any opportunity to change decisions which they have made in their life. Majority of the respondents said they would take the opportunity, with a very small number saying they would not change anything. Also when asked how
they handled stress, they gave solutions like religion, help from family and friends and mainly amongst women, shopping.

2.6.2 Review of Technical Approach of Related Work

Previous research reviewed all had similar goal which was to identify the factors which influenced the level of happiness of students. Generally the research method employed was through the use of surveys to gather data. This was done through the use of questionnaires or interview methods. In order to carry out data analysis, statistical methods such as standard deviation, mean and standard error were the most frequently used method.

The ordered probit regression method was the most notable technical method used (Chan et al., 2005). In the study carried out by Chan et al (2005) of students in the University of Western Australia the ordered probit model was used. This model is suitable for categorical data which have an underlying ordering, and difference between dependent variable categories do not have the signify distance (McKelvey and Zavoina, 1975).

Linear relationship is the basis of the ordered probit model; the equation is given by Chan et al (2005) as below.

\[ SA_i = X_i \beta + \epsilon_i \]

Where \( SA_i \) is the satisfaction scale which is unobserved, \( \epsilon_i \) is the error and assumed to follow a normal distribution and \( \beta \) is the vector of parameters to be used for estimation and \( X_i \) is the vector of explanatory independent variables (Chan et al., 2005). The vector includes gender, student monetary resources, age, satisfaction with school work, university resources and environments, health, reputation of university, relationships formed with other students, getting a part time job and time management. The statistical analysis was based on the SA1, the observed indicator given by the statement “Overall I am happy with my university life” (Chan et al., 2005). The value of SA1 was gotten from the responses based on a scale of five with 5 = “strongly agree” and at the other extreme 1 = “strongly disagree”. At the end of the questionnaire a similar question represented by SA2 with same scale of rating was used to check the reliability of the question. SA1 (observed scale) for an individual i, which was linked to SA (unobserved variable) as depicted in the equation by chan et al (2005) below:

\[
\begin{align*}
SA_{1i} &= 0 \text{ if } SA_1 \leq \mu_0 (=0) \\
SA_{1i} &= 1 \text{ if } \mu_0 \leq SA_1 < \mu_1 \\
SA_{1i} &= 2 \text{ if } \mu_1 \leq SA_1 < \mu_2 \\
SA_{1i} &= 3 \text{ if } \mu_2 \leq SA_1 < \mu_3 \\
SA_{1i} &= 4 \text{ if } \mu_3 \leq SA_1
\end{align*}
\]
Where $\mu_s$ is the ‘unknown threshold parameter’ which separates the adjacent groups (Chan et al., 2005). The value of the first threshold parameter was normalised to zero. The resulting probabilities were calculated with the normal distribution which was given by Chan et al (2005) as:

$$\text{Prob} (SA_1=j) = \Phi(\mu_j - \chi \beta) - \Phi(\mu_{j-1} - \chi \beta),$$

Where the ‘standardised normal cumulative distributive function’ is given as $\Phi$. (Chan et al., 2005). The results gotten from the model, shows that a variable has a higher probability of impacting $SA_1$ (overall satisfaction) if a variable is preceded by a positive coefficient. The reverse is the case if the variable is preceded by a negative coefficient. The level of importance and independence of student’s response to the satisfaction question was examined using the chi-squared statistical method. How well the observable variable is able to the satisfaction of student with university life is indicated by the ordered probit model’s Likelihood Ratio Index.

Each observed variable’s coefficient showed the extent to which a specific variable impacted $SA_1$ using the t statistics. The probability and percentage change which would happen following a rise or reduction of an independent variable by an element was shown using the marginal effect. The results gotten from the marginal effects showed which independent value will improve student’s happiness significant and those that would not. The probability predicted showed which variable had the most impact on student’s distribution. It equally showed how distribution would occur depending on the responses gotten from the five categories in $SA_1$.

The study carried out by Mangeloja and Hirvonen (2007) with students of the University of Jyvaskyla also made use of the ordered probit model. The analysis performed was similar to that done by Chan et al (2005) (Mangeloja and Hirvonen, 2007). The results gotten from the analysis was found to be consistent with that gotten by Chan et al (2005). The main difference was that the key factors which impacts the happiness of students in university of Western Australia was not as important in the University of Jyvaskyla.

The sample size used by Al-Naggar et al (2010) was small; hence he used a basic approach in performing his analysis. Group discussion was used as a method of data collection, and based on the response he manually analysed the data.

In the University of Algarve interview method was used for data collection and each student was asked to list activities which made them happy (Guerreiro and Vieira, 2007). Figure 1 shows the percentage of happiness reported by students in the University of Algarve.
SPSS and Microsoft excel were the tool used for data analysis. Standard error, standard deviation, mean, level of significance and percentage analysis were the main statistical methods employed in the analysis. When analysing the data by gender the variables were plotted against each other and the variable with the highest impact on the happiness and life objective of students was shown.

Tables, bar charts and histograms were also used to analyse the happiness of students. A new dimension added to test data reliability was the added question to compare one’s happiness with the happiness of others.

2.6.3 Weakness and Limitations of Reviewed Works
Studies which were reviewed in this section motivated and gave direction to the study carried out in this project. However there were weakness and limitations which were noted. The previous studies
were limited to identifying factors which impacted student’s happiness. Models for predicting the future level of student’s happiness was not considered.

The ordered probit approach (Chan et al., 2005; Mangeloja and Hirvonen, 2007) only built models to show what factors affects the happiness level of students. Thought was not given to build models that would be able to model the relationship between the independent variables and future happiness of students. A major limitation of this approach is that it is suited for only categorical data hence would not be able to handle continuous or numeric data. Also the output gotten from this approach is complex and difficult to interpret.

The method which was used to identify factors which impacted student’s happiness was not clear. As a result we are not able to ascertain if the identified factors are a true reflection from the perspective of students. Students were only able to give response based on the list of factors the researcher felt was relevant. Another weakness of previous studies is the absence of any check to validate student’s responses. Most times students might feel that certain factors impact their level of happiness, while in reality that is not true.

A major limitation of the studies is the localisation of data gotten. Data used for analysis were only gotten from within the same place, and had no global reach; hence segregation based on cultural difference was not performed.

2.7 Background Summary
This section covered the background study of subjective well-being and happiness, and its roots in psychology and economics. Measures of happiness and definition of happiness was established. Previous studies on student’s happiness were reviewed along with the methodology used; limitation and weaknesses.

The use of machine learning techniques have grown and has recorded lots of success in non-computing fields such as medical diagnosis, stock market analysis, credit card fraud detection, gene detection and medical imaging. Machine learning techniques are capable of building predictive models and spotting complex relationships in data. This study is interdisciplinary and should not be confused as being a social science project due to the nature of the background study.

Statistical methods have been applied successfully in previous studies of happiness economics. The capabilities of machine learning would raise the standard and add another dimension by offering a higher dimension of data analysis.
Chapter 3: Project Progress

3.1 Research Methodology
As described in chapter one, this project aims to identify the factor which influences student’s happiness and to build a mathematical model of student’s happiness. By employing machine learning methods and computational intelligence to show the relationship between the factors and the level of happiness of students as well as the extent to which each factor plays a role in student’s happiness.

The research process is descriptive, analytical and predictive and is conducted from a quantitative point of view. The method designed to perform this research is in line with the project objectives listed in section one. The steps needed to carry out this research are listed in the project plan and are shown below:

<table>
<thead>
<tr>
<th>Step</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Background Research</strong></td>
</tr>
<tr>
<td>Review of previous studies is carried out, to gain a broader understanding of the project.</td>
</tr>
<tr>
<td><strong>Identification of Happiness Factors</strong></td>
</tr>
<tr>
<td>For analysis, the factors which influences the level of happiness of students need to be identified</td>
</tr>
<tr>
<td><strong>Questionnaire Design</strong></td>
</tr>
<tr>
<td>Based on the identified factors, a questionnaire would be designed</td>
</tr>
<tr>
<td><strong>Data Collection</strong></td>
</tr>
<tr>
<td>Data would be collected which would form the dataset which would be used for analysis</td>
</tr>
<tr>
<td><strong>Identify Machine Learning Technique</strong></td>
</tr>
<tr>
<td>Machine learning techniques which would be used for analysis are studied and identified</td>
</tr>
<tr>
<td><strong>Identify Data Analysis Toolbox</strong></td>
</tr>
<tr>
<td>The toolbox which would be used to create the model would be identified</td>
</tr>
<tr>
<td><strong>Experiment Design</strong></td>
</tr>
<tr>
<td>For a successful experiment the experiment design would be planned carefully</td>
</tr>
<tr>
<td><strong>Data Pre-processing</strong></td>
</tr>
<tr>
<td>Data cleaning, transformation and integration is done to ensure data reliability and consistency</td>
</tr>
<tr>
<td><strong>Perform Experiments</strong></td>
</tr>
<tr>
<td>Experiments would be performed and model generated</td>
</tr>
<tr>
<td><strong>Result Analysis and Evaluation</strong></td>
</tr>
<tr>
<td>The best model is selected and analysis and evaluation is conducted</td>
</tr>
</tbody>
</table>

Figure 3: Research Methodology
3.2 Evaluation Plan

In order to determine if the stated objectives are achieved and also measure the success of the project an evaluation needs to be carried out at every stage of the project. One of the advantages of performing evaluation while project is on-going is to serve as a check ensuring it meets required standard and objectives. Formative evaluation would be used in the course of the project. University research specifications would be used to evaluate if the required standard is followed during the project. The objectives, strategies and method of evaluation are listed in the table below:

Table 1: Evaluation Plan

<table>
<thead>
<tr>
<th>Objectives</th>
<th>Strategy</th>
<th>Method of Evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identification and analysis of factors which impacts the level of student’s happiness.</td>
<td>Analytical Research would be carried put.</td>
<td>Results of analysis and identified factors would be contained in dissertation report.</td>
</tr>
<tr>
<td>Dataset used for analysis. This would be gotten from student’s response to the questionnaire on happiness.</td>
<td>Data gotten through the use of surveys such as questionnaires.</td>
<td>Dataset with a sample size of more than 300.</td>
</tr>
<tr>
<td>Identification of suitable machine learning techniques to be used for learning.</td>
<td>Various machine learning techniques would be studied.</td>
<td>Clear description and identification of suitable machine learning methods selected.</td>
</tr>
<tr>
<td>Learn from the data and develop the mathematical relationship to model the student’s happiness.</td>
<td>Perform experiments through the use of machine learning techniques.</td>
<td>Evaluation of accuracy and fit of the predictive model. Using cross validation.</td>
</tr>
<tr>
<td>Factors which impact the level of student’s happiness.</td>
<td>Analysis of results gotten from the experiments</td>
<td>Comprehensive report showing analysis and understanding of results.</td>
</tr>
</tbody>
</table>

3.3 Collaborative and Communication tools

For a successful project tools for communication and collaboration need to be identified and used. This tools which would ensure seamless communication between all involved parties and help in successful completion of project. The tools identified are listed below


**Emails:** Email would serve as the main tool of communication with my project supervisor. This would enable quick review of work, feedback and meeting scheduling. It would also assist in sending out web based survey to target participants.

**Web based Survey:** A questionnaire would be developed online through survey tools such as Google docs and Survey monkey. This would enable participation of students from different locations in the project, enabling collection of data from diverse individuals. The data collected would aid in a comprehensive analysis.

**Facebook:** Facebook is an online social network tool with a large number of users. This tool would aid in distribution of survey to students in different parts of the world. Links to the survey would be put up in students discussion groups joined other student’s activities ensuring participation from a large number.

### 3.4 Project Plan
Good planning is one of the pre-requisite of a successful project. A good methodology is not the only factor to ensure that project is successful and completed on time. A project plan was developed which would be used in the course of the project to keep updated with progress made as well as ensure proper time management. **Appendix 1** shows the project plan which would be used in this project. The list of project milestones and deliverables are shown below.

#### 3.4.1 List of Project Milestones
1) Completion of data collection (due 7th June 2013)
2) Submission of progress report (due 10th May 2013)
3) Completion of analysis and evaluation of results (due 31st July 2013)
4) Submission of final dissertation (due 6th of September 2013)

#### 3.4.2 List of Project Deliverables
- Factors impacting the students happiness
- Datasets formed from the questionnaire collected
- Experiment results from different computational intelligence and machine learning methods
- Best Machine Learning Methods recognised
- Mathematical model of student’s happiness
- Analysis and Recommendation based on the model
- Completed project Report
3.5 Progress Report

At the point of writing the preliminary report in May 2013 the progress made so far in the dissertation project is presented in this section. Further work would commence immediately my last exam is completed. The project plan showed in Appendix 1 shows the status of each tasks. Details of the progress made are described as follows.

Background Research

Extensive background research has been conducted. This is reflected in the literature review section included in this report. The section covered the background study of subjective wellbeing and happiness, and its roots in psychology and economics. Measures of happiness and definition of happiness was established. Previous studies on student’s happiness were reviewed along with the methodology used; limitation and weaknesses. The nature of the background research should not be confused for a social science project as this is an interdisciplinary project. The use of machine learning techniques cuts across all fields computing and non-computing alike. Some examples of applications of machine learning techniques in non-computing area include but are not limited to fraud detection in banking, medical diagnosis and imaging in medicine, and stock market analysis in the world of investments.

Factors Influencing Student’s Happiness

The factors which impact the happiness level of students have been identified. This was done through several discussions and interviews with students. A lot of students came up with a list of factors which they considered to impact their level of happiness. Some other important factors were considered while reviewing previous research carried out in the past by Chan et al (2005), Mangeloja and Hirvonen (2007) and Guerreiro and Vieira (2007). This factors combined were further analysed and evaluated and with the help of my supervisor I was able to organise it to 11 factors. These factors would serve as features during data analysis. The selected factors which were identified are listed below:

- Interesting courses
- Academic performance
- Effectiveness of lectures and courses
- University facilities
- Extracurricular activities
- Financial situation
- Relationships (family and partner)
- Living conditions
- Relationship formed with other students
• Future job prospects
• Existing Visa policy (for international students only)

Questionnaire

Based on the identified factors which impact the student’s happiness level, a questionnaire was designed. The questionnaire was designed to be reflective of its intended use, simple, understandable and easily completed. It included only closed ended questions (limited set of answers) to enable simplicity. The questionnaire was divided into two sections. The first section contained demographic information of the student such as age, gender, nationality, university, faculty and degree. In the second section students were required to rate their current level of happiness on a scale of one to ten, data gotten from this would serve as the dependent variable. Students were also required to assess how the listed factors impact their happiness level and also rate their current state of happiness. This was rated on a scale of one to ten as well and served as the independent variable. The design of questionnaire was in two forms, online through the use of surveys and paper based. The paper based questionnaire designed is shown in Appendix 2, it was limited to one page so as to capture the interest of students. This link shows the Web based Survey that was developed.

Data Collection

Data collected would form the dataset which would be used for analysis. The subjects selected to participate in the survey would be restricted to university students. There would be no restriction as to university location or nationalities of participants. The aim of the data collection is to get as much data as possible from diverse students in different universities. This would further aid in performing a comprehensive analysis. Web based survey using Google docs and paper based questionnaire where the primary method of data collection used. Currently data collection is still in progress and is limited to students who are currently enrolled in a University. So far a total of 278 responses have been gotten. Preliminary analysis shows that 65% of the respondents are male and 35% are females, 52% of respondents are currently enrolled in a postgraduate course, Majority of the respondents are in the faculty of engineering and physical science also the average reported overall level of happiness is rated to be 7.

Machine Learning Techniques

Machine learning methods are used to perform data analysis the advantage of this technique is that it helps in building predictive models as well as spotting complex relationships. The identified machine learning methods which would be used for analysis would be briefly discussed here.
Support vector machine (SVM)

The support vector machine was first proposed by Vapnik (1995). The Support vector Machine is built on the basis of ‘structured risk minimization’ (Pai and Lin, 2005). SVM can be used for both regression and classification problems with a high rate of accuracy. Regression involves predicting or estimation a response, the output takes continuous variable. In Classification the output value takes the class label which is used for identifying group membership. When using the support vector machine the function is given as

$$y = f(x_1, x_2, x_3, ..., x_{11})$$

Where y is the Happiness index (on a scale of 1 to 10) of the student and $$(x_1, x_2, x_3, ..., x_{11})$$ is the features which are the factors which affect the student’s happiness level.

SVMs occur as a result of theoretical bounds of the generalisation error. The upper bound of the generalisation error is not dependent on the dimensionality of the space. The SVM regression function is given by Pai and Lin (2005) as shown below

$$y = w\phi(x) + b,$$

Where $$\phi(x)$$ is used as the feature. Pai and Lin (2005) shows that w and b coefficients are gotten by minimizing

$$R(C) = C\frac{1}{N} \sum_{i=1}^{N} L_\varepsilon(d_i, y_i) + \frac{1}{2} ||w||^2,$$

$$L_\varepsilon(d, y) = \begin{cases} |d - y| - \varepsilon & |d - y| \leq \varepsilon, \\ 0 & \text{others}, \end{cases}$$

‘$$L_\varepsilon(d, y)$$’ is the ‘$$\varepsilon$$ intensive loss function’ (Pai and Lin, 2005) which shows that errors below the slack variable $$\varepsilon$$ are not penalized. ‘$$\frac{1}{2}||w||^2$$’ is used to maximise the margin and ‘$$C\frac{1}{N} \sum_{i=1}^{N} L_\varepsilon(d, y)$$’ is used to constrain the optimisation to ensure correct classification (Pai and Lin, 2005). The trade-off between allowing slack and optimising the loss function is assessed using C, hence minimizing the error function subject to the constraints.

Introducing the Lagrangian multipliers ($$\alpha_i$$ and $$\alpha_i^*$$) and maximising the dual function results in the below equation which satisfies the equalities (Pai and Lin 2005, eq 14)

$$f(x, \alpha, \alpha^*) = \sum_{i=1}^{l} (\alpha_i - \alpha_i^*) K(x, x_i) + b.$$
\( K(x, x_i) \) is the kernel function. The Gaussian kernel \( K(x, x_i) = \exp^{-\|x-x_i\|^2/2\sigma^2} \) would be employed in developing the model; this is because the Gaussian kernel is known to have a good performance (Pai and Lin, 2005).

SVM is selected as one of the machine learning technique employed for model generation. This is because of its reported high accuracy, fast rate of classification and its ability to handle non-linear and categorical data. Despite its numerous advantages some of the disadvantages of this technique is it is not easy to use due to its parameter tuning. Another disadvantage is that support vector machine’s black box nature, makes results gotten difficult to interpret.

**Regression**

Regression analysis is mostly used for forecasting and numeric prediction, it is a statistical method used for estimating relationships among variables. Regression is done by establishing the relationship between the dependent variable and the independent variables. Regression analysis helps in exploring the relationship and understanding how the dependent variable changes when there is a change in any one of the independent variable and the other independent variable is kept constant. The independent variable associated closely to the dependent variable is used to develop the predictive model. Given a data set \( \{y_i, x_{i1}, x_{i2}, x_{i3}, \ldots, x_{in} \} \) the relationship between the dependent and independent variable is given by the linear equation

\[
y_i = b_1 x_{i1} + b_2 x_{i2} + b_3 x_{i3} + \ldots + b_n x_{in} + e_i
\]

where \( y_i \) is the dependent variable (happiness index), \( x_{i1}, x_{i2}, x_{i3}, \ldots, x_{in} \) is the independent variable (factors which impacts the student happiness), \( b \) is the regression coefficient which shows the rate of change of the dependent variable as a function of changes in the independent variable and \( e_i \) is the error which captures all factors which impacts the dependent variable.

The predictive power and accuracy of the model produced are improved by minimizing the error. Linear regression does not have a high accuracy when compared with the support vector machine. It would be used in this project as the base on which other methods would be compared.

**Decision tree**

Decision tree is a machine learning method used to build predictive models, based on several input variables. It is the combination of computational and mathematical technique to help in the categorisation, generalisation and description of a given dataset. The learning algorithm employed by this method to build tree from data is the recursive algorithm. Decision tree learning is used in this project because it is capable of handling both categorical and numerical data, its ability to handle large
amount of data in a sort time and most importantly results are simple to interpret and understand. An example of a decision tree is given in the figure 4 below

![Decision Tree Example](image)

**Figure 4: Decision tree example (Witten and Frank, 2005; Han and Kamber, 2006; M.O.N.K, n.d.)**

From the example above the decision to play or not is based on weather condition. The visual representation enables easy interpretation of the results gotten. Weather condition can also be classified in terms of numeric data for example 15°C can be used to depict temperature. The model produced by the decision tree can visually show how each factors affects the student’s happiness and is capable of handling numeric data.

**Clustering**

Clustering is the process of grouping similar set of data together (clusters). It is a common technique for statistical data analysis and it is used in machine learning. The process of clustering would be employed in the project. This would enable proper segregation and analysis of data, various machine learning methods would be applied to each identified cluster to form a model and proper understanding of the dataset.

**Feature Selection**

Feature selection aids in developing good learning models, it takes subset of relevant feature from a pool of features for use in creating model. Feature selection would be employed in this project to reduce computation cost, improve accuracy, improve data/model understanding, and reduce cost of future measurements. There are two methods of performing feature selection which are filter and wrapper selection. Filter selection selects features independently of how well the machine learning technique performs and is generally a pre-processing step. Wrapper selection is dependent on the
performance of the machine learning technique (model dependent) and is usually more computationally intensive than the filter method.

**Cross Validation**

Cross validation is a method for assessing the predictive power of a learning technique. It estimates how accurately a predictive model will perform when put to use. This method can also be used to compare the performance of different machine learning methods as well as in variable selection, and would be used in the project.

Other techniques considered are Naive Bayes which is a special type of Bayesian Network and neural network. But due to the low level of accuracy associated with Naive Bayes as well as the low classification speed and danger of over fitting associated with neural network these were dropped.

**Project Toolbox**

The identified project toolbox which would be used for the project are listed below

**WEKA:** This is a free machine learning software suite developed in java. It provides a comprehensive collection of data modelling, data pre-processing and machine learning techniques. Its graphical user interface enables easy implementation and usage.

**IBM SPSS:** This software package is used for statistical analysis, data mining and text analytics.

**MATLAB:** This is a numerical computing environment that can be used for data visualization, analysis and computation. MATLAB provides advantage in solving technical or numerical computational problems much faster than regular programming language. This is as a result of being a scientific programming language.

**Experiment Design**

In order to achieve the aim of the project, a sequence of experiments has to be thought out and designed. Figure 5 below shows the summarised view of the method of experimentation. Further details would be given below
The process starts with data pre-processing which involves data cleaning, integration and transformation techniques so as to remove inconsistent data as well as detect outliers and anomaly. After data pre-processing have been done, an overall analysis would be done on the data to get a general understanding of the model. In order to get the best possible results, clustering would be performed on the dataset. This would enable grouping of the data sets into similar set which would result in better understanding of the behaviour of the datasets and also produce a better mathematical model. Clustering would be performed using the K-nearest neighbour technique and also intuitively using demographic information such as gender, age, nationality etc. Model generation would be performed on the different clusters formed using the identified machine learning techniques such as support vector, decision tree and regression. When using the k-nearest neighbour technique only groups with a large number of data points would be used to perform the experiment. After the models have been generated using the various machine learning techniques. The models would be compared using the specified evaluation metrics; this would ensure the best model is used as a representation of the student’s happiness.

Feature selection would also be employed in each cluster in order to improve the predictive ability of the model. Irrelevant and redundant subsets would be filtered out and only the subsets which improve the model’s predictive ability would be retained. The various machine learning techniques would be applied to clusters in which feature selection have been performed to see if it improves or decreases the model performance. It would also be used to pick out the factors which have the highest impact on the level of happiness of student’s by using the ranking allocated by the evaluator.
To select the best model, the dataset would be split into training and testing set. The accuracy of the models would be determined by how accurately the test data is predicted by the model after training. To determine the accuracy of the models error metrics would be used. The model with the best accuracy would be used as the final model of the student’s happiness. This model would also show how each factor influences the level of student’s happiness.

3.6 Project Progress Summary
This chapter shows the research methodology to be used in the course of the project. The evaluation plan and project plan were also listed. The progress made in the project so far was discussed. Presently background research has been carried out and the research methodology to be used has been developed. The factors which influence the level of student’s happiness has been identified and used to create the questionnaire which would be used to gather data for analysis. The machine learning techniques and toolbox which would be used for analysis was identified and experiment design developed. This chapter also showed the project plan and evaluation plan.
References


Easterlin, R.A. (1995) 'Will raising the incomes of all increase the happiness of all?', *Easterlin, Richard A. "Will raising the incomes of all increase the happiness of all?" Journal of Economic Behavior & Organization*, vol. 27, no. 1, pp. 35-47.


Fredrickson, B.L. (1998) 'What good are positive emotions?', Review of general psychology: Journal of Division 1, of the American Psychological Association, vol. 2, no. 3.


Guerreiro, M. and Vieira, G. (2007) 'Students' happiness and well-being at the University of Algarve, Portugal.'.


Likert, R. (1932) 'A technique for the measurement of attitudes', *Archives of psychology*.


## APPENDICES

### Appendix 1: Project Plan

<table>
<thead>
<tr>
<th>ID</th>
<th>Task Name</th>
<th>Duration</th>
<th>Start Date</th>
<th>End Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Learning the student's happiness model</td>
<td>157 days</td>
<td>31/01</td>
<td>06/06</td>
</tr>
<tr>
<td>2</td>
<td>Background Research</td>
<td>107 days</td>
<td>31/01</td>
<td>15/06</td>
</tr>
<tr>
<td>3</td>
<td>Background Reading</td>
<td>10 days</td>
<td>31/01</td>
<td>10/06</td>
</tr>
<tr>
<td>4</td>
<td>Literature Review</td>
<td>28 days</td>
<td>31/01</td>
<td>28/05</td>
</tr>
<tr>
<td>5</td>
<td>Identification of Happiness Factor</td>
<td>6 days</td>
<td>31/01</td>
<td>07/05</td>
</tr>
<tr>
<td>6</td>
<td>Questionnaire Design</td>
<td>6 days</td>
<td>31/01</td>
<td>07/05</td>
</tr>
<tr>
<td>7</td>
<td>Preliminary Report Write Up and Submission</td>
<td>7 days</td>
<td>31/01</td>
<td>07/05</td>
</tr>
<tr>
<td>8</td>
<td>Data Collection</td>
<td>82 days</td>
<td>31/01</td>
<td>02/06</td>
</tr>
<tr>
<td>9</td>
<td>Milestone 2 - Data Collection Completed</td>
<td>0 days</td>
<td>31/01</td>
<td>02/06</td>
</tr>
<tr>
<td>10</td>
<td>Methodology Development</td>
<td>47 days</td>
<td>31/01</td>
<td>09/05</td>
</tr>
<tr>
<td>11</td>
<td>Identify and Develop Methods</td>
<td>11 days</td>
<td>31/01</td>
<td>01/06</td>
</tr>
<tr>
<td>12</td>
<td>Search and Identify Data Analysis tool</td>
<td>11 days</td>
<td>31/01</td>
<td>01/06</td>
</tr>
<tr>
<td>13</td>
<td>Experiment Design</td>
<td>7 days</td>
<td>31/01</td>
<td>07/05</td>
</tr>
<tr>
<td>14</td>
<td>Progress Report Write up</td>
<td>34 days</td>
<td>31/01</td>
<td>04/06</td>
</tr>
<tr>
<td>15</td>
<td>Review and Submission of Progress report</td>
<td>11 days</td>
<td>31/01</td>
<td>02/06</td>
</tr>
<tr>
<td>16</td>
<td>Milestone 1 - Progress Report Submitted</td>
<td>0 days</td>
<td>31/01</td>
<td>02/06</td>
</tr>
<tr>
<td>17</td>
<td>Experiment, Data Analysis and Evaluation</td>
<td>37 days</td>
<td>31/01</td>
<td>04/06</td>
</tr>
<tr>
<td>18</td>
<td>Data Preprocessing</td>
<td>6 days</td>
<td>31/01</td>
<td>07/05</td>
</tr>
<tr>
<td>19</td>
<td>Experiments</td>
<td>11 days</td>
<td>31/01</td>
<td>01/06</td>
</tr>
<tr>
<td>20</td>
<td>Result Analysis and Evaluation</td>
<td>20 days</td>
<td>31/01</td>
<td>11/06</td>
</tr>
<tr>
<td>21</td>
<td>Milestone 3 - Result Evaluation Completed</td>
<td>0 days</td>
<td>31/01</td>
<td>11/06</td>
</tr>
<tr>
<td>22</td>
<td>Final Dissertation Report</td>
<td>59 days</td>
<td>31/01</td>
<td>10/06</td>
</tr>
<tr>
<td>23</td>
<td>Dissertation Write up</td>
<td>47 days</td>
<td>31/01</td>
<td>04/06</td>
</tr>
<tr>
<td>24</td>
<td>Review and Submission of Dissertation</td>
<td>12 days</td>
<td>31/01</td>
<td>03/06</td>
</tr>
<tr>
<td>25</td>
<td>Milestone 4 - Dissertation Submitted</td>
<td>0 days</td>
<td>31/01</td>
<td>03/06</td>
</tr>
</tbody>
</table>

*Completed*

*In progress*
BASIC INFORMATION (Please tick where appropriate)

Gender: □ Female  □ Male
Nationality: □ UK/EU  □ Indian  □ Nigerian  □ Chinese  □ Other (Please state) ______________
University: □ University of Manchester  □ Other UK Universities  □ Nigerian Universities  □ Others
Faculty: □ Humanities  □ Engineering & Physical Sciences  □ Medicine & Human Sciences  □ Life Sciences
Degree: □ Undergraduate  □ Postgraduate  □ Final Year

QUESTIONNAIRE INSTRUCTIONS

For Your View Please rate what effect the following factors have on your level of happiness on a scale of 1 – 10 with 10 being the strongest influence and 1 being the least influence.

Under Current Status, Please grade your current condition of each factor on a scale of 1 to 10 with 10 being the most satisfactory and 1 being the least satisfactory.

As a student, please rate what effects the following factors have on your level of happiness. Please indicate your current condition for each factor. (Please tick where appropriate)

Please rate your current state of happiness (on a scale of 1 to 10): □ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7 □ 8 □ 9 □ 10

<table>
<thead>
<tr>
<th>Your View (Scale of 1 – 10)</th>
<th>Current Status (Scale of 1 – 10)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1  2  3  4  5  6  7  8  9  10</td>
<td>1  2  3  4  5  6  7  8  9  10</td>
</tr>
<tr>
<td>How interesting are the courses you take</td>
<td>How well you perform in your studies</td>
</tr>
<tr>
<td>1  2  3  4  5  6  7  8  9  10</td>
<td>1  2  3  4  5  6  7  8  9  10</td>
</tr>
<tr>
<td>Effectiveness of lectures and course structure</td>
<td>Effectiveness of lectures and course structure</td>
</tr>
<tr>
<td>1  2  3  4  5  6  7  8  9  10</td>
<td>1  2  3  4  5  6  7  8  9  10</td>
</tr>
<tr>
<td>How the available University Facilities meet your Learning needs</td>
<td>How the available University Facilities meet your Learning needs</td>
</tr>
<tr>
<td>1  2  3  4  5  6  7  8  9  10</td>
<td>1  2  3  4  5  6  7  8  9  10</td>
</tr>
<tr>
<td>Available extracurricular activities (Leisure time, Sports and Social Activities)</td>
<td>Available extracurricular activities (Leisure time, Sports and Social Activities)</td>
</tr>
<tr>
<td>1  2  3  4  5  6  7  8  9  10</td>
<td>1  2  3  4  5  6  7  8  9  10</td>
</tr>
<tr>
<td>Financial Situation Stability</td>
<td>Financial Situation Stability</td>
</tr>
<tr>
<td>1  2  3  4  5  6  7  8  9  10</td>
<td>1  2  3  4  5  6  7  8  9  10</td>
</tr>
<tr>
<td>Well balanced Family and Partner relationship</td>
<td>Well balanced Family and Partner relationship</td>
</tr>
<tr>
<td>1  2  3  4  5  6  7  8  9  10</td>
<td>1  2  3  4  5  6  7  8  9  10</td>
</tr>
<tr>
<td>How Comfortable your Living Conditions are</td>
<td>How Comfortable your Living Conditions are</td>
</tr>
<tr>
<td>1  2  3  4  5  6  7  8  9  10</td>
<td>1  2  3  4  5  6  7  8  9  10</td>
</tr>
<tr>
<td>Relationships formed with other students</td>
<td>Relationships formed with other students</td>
</tr>
<tr>
<td>1  2  3  4  5  6  7  8  9  10</td>
<td>1  2  3  4  5  6  7  8  9  10</td>
</tr>
<tr>
<td>Getting a good Job after Graduation (Future job prospects)</td>
<td>Getting a good Job after Graduation (Future job prospects)</td>
</tr>
<tr>
<td>1  2  3  4  5  6  7  8  9  10</td>
<td>1  2  3  4  5  6  7  8  9  10</td>
</tr>
<tr>
<td>*Existing policy on Post study Visa (International Students only)</td>
<td>*Existing policy on Post study Visa (International Students only)</td>
</tr>
</tbody>
</table>

Appendix 2: Questionnaire Design