**AUTOMATED CAREER GUIDANCE EXPERT SYSTEM USING CASE-BASED REASONING**

**BY**

**JEROME,OTOBONG BASSEY**

**11/EG/CO/532**

**SUBMITTED TO**

**THE DEPARTMENT OF ELECTRICAL/ELECTRONICS AND COMPUTER ENGINEERING**

**FACULTY OF ENGINEERING**

**UNIVERSITY OF UYO**

**IN PARTIAL FULFILLMENT OF THE REQUIREMENT FOR THE AWARD OF BACHELOR OF ENGINEERING (B.Eng) DEGREE IN COMPUTER ENGINEERING**

**SEPTEMBER 2016**

TABLE OF CONTENT

Chapter 1

Introduction

1.0 BACKGROUND STATEMENT

1.1 STATEMENTS OF PROBLEM

1.2 OBJECTIVE OF STUDY

1.3 SCOPE AND LIMITATION

CHAPTER TWO

2.0 LITERATURE REVIEW

CHAPTER THREE

3.0 Materials and Methods

3.1 Requirement Gathering

3.1.1 Data Collection

3.1.2 Data pre-processing

3.2 Planning

3.2.1 Rule-Based expert systems

3.2.2Neural Networks

3.2.3 Case-Based Reasoning

3.3 Design

3.4 CODING

3.5 THE FIRST DELIVERABLE

**Chapter 1**

**Introduction**

**1.0 BACKGROUND STATEMENT** Career guidance refers to an organized system of social and professional work on providing informed suggestions or advice to an individual to aid the individual’s choice of direction and orientation. This is often in education or other professional activities, with the goal of achieving a professional identity, in accordance with personal traits and the labor market demand for specific occupations. Furthermore, it plays a key role in helping labor markets, work and education systems meet their goals. The concept of career guidance has been widely used in the past and in several varieties. According to Herr and Cramer(1996), “It is largely a verbal process which involves the counselor and counsel-lee (s) developing a dynamic and collaborative relationship, focused on identifying and acting on the counselee’s goals, in which the counselor employs a repertoire of diverse techniques and processes, to help bring about self-understanding, understanding of behavioral options available, and informed decision making in the counsel lee, who has the responsibility for his or her own actions”.

Remarkably, guidance and counseling is an integral practice in schools and colleges, the intent of which is to provide pieces of advice to students in order to enable them improve in their academic performances, choose a befitting career and maximize their individual potentials. However, the ubiquity of the Information and Communication Technology (ICT) has changed the way this activity is carried out in most developed countries. Motivated by the aforesaid, the thrust of this project is to automate the process of career guidance using an Artificial Intelligence approach, specifically known as Case-based reasoning (CBR).

CBR is a general artificial intelligence paradigm for reasoning from experience. It is a natural way of solving problems that are similar to previously solved problems. This technique is employed by lawyers when they refer to previous case files in order to handle a current case which is similar to it. Moreover, other professionals such as medical doctors, engineers, etc. often use CBR in their daily activities. The idea behind this technique is to have a memory base of all cases experienced and their corresponding solutions such that if a new case is encountered, a similarity metric is used to compare its similarities with the cases in the memory base and the solution to the most related cases is applied to the new case. Some of the similarity metrics often used include; Euclidean Norm, Manhattan distance, Hamming Distance and so on. CBR is also known as Instance-based Reasoning, K-Nearest Neighbor algorithm (KNN), or memory-based learning.

This project is an expert-based software system often referred to as expert system tailored to mimic the intelligence of a human academic counselor in advising and guiding senior secondary school’s students in choosing courses for institution of higher learning. Noticeably, rule-based expert systems are the most popular expert systems nowadays. This type of expert systems makes use of production or IF-THEN rules stereotyped to the knowledge acquired from the expert by the knowledge engineer or the system developer. However, due to knowledge acquisition bottleneck, the need for a data-driven expert system development technique becomes essential. This bottleneck culminates, among many reasons, from the inability of most experts to explain what they know and how they know them; scarcity of experts; and ambiguity due to the expert's diction. Furthermore, in data-driven expert system development approach, a machine learning algorithm is used to glean rules buried in data generated by a human expert without having to develop the rules explicitly. These techniques include; Neural expert systems, Memory-based expert system, Bayesian expert system, and so on. In this project the memory-based expert system is used due to its ease of implementation compared to other techniques. An excel dataset in .xls format containing one thousand (1000) different cases with eight (8) features and one output. The features consist of important subjects offered in senior secondary schools. Without any loss of generality, the subjects includes mathematics, English language, physics, chemistry, biology, geography, Literature, accounting, Christian Religious Knowledge(CRK), and Government, for each case the output contains a human counselor's course recommendation based on the students performances in the eight subjects. Using this dataset as a memory-base, a GUI- based MATLAB application is developed to meet the experience of the human counselor captured in the dataset to recommend courses to a new user.

**1.1 STATEMENTS OF PROBLEM**

The lack of quality career guidance in Nigerian secondary schools, caused by shortage of human and time resources that the process demands, has led to various problems such as;

* The choice of unsuitable careers resulting in widespread poor performance at the industry.
* A mismatch between the careers most secondary school graduates choose and the ones of which they have natural aptitude for.
* Career- related confusions due to student lack of an accessible, easy-to-relate-to and trustworthy career decision support resource at their disposal.

These problems which are mostly due to the fact that;

* At age of 17, most students are not mature enough to know precisely what career to follow; they are not sufficiently aware of what goes on in a particular area and which academic majors are associated with their areas of interests.
* Human experts advice is not available all the time, most secondary school guidance counselors are assigned other class duties which makes it difficult for them to spend as much time providing guidance services.
* Knowledge of each students acquired by the counselor staff cannot be transferred to other staff in cases where there is a change of the school’s guidance counselor.
* Teachers sometimes get too emotionally involved and are unable to make rational decision for the students.
* The counseling staff sometimes gets overworked and tired reducing their efficiency towards guidance and counseling of the students.
* Retrieving past information about the students can be a rigorous process.

These are some of the problems which the expert system approach aims to take care of.

**1.2 OBJECTIVE OF STUDY**

Guidance and counseling is an integral and requisite part of any education system and a lot of emphasis should be laid on its implementation to ensure that its roles and objectives are realized. Guidance and counseling provides information in three critical areas: academic, personal/social, and career. In secondary school system, this information can help students resolve emotional, social or behavioral problems and help them develop a clearer focus or sense of direction. The selection of career paths for students after A-levels/intermediate is an attention requiring concern. This paper presents an automated system that mimics a one-to-one meeting with a professional career counselor. The system supports people in developing their own career opting competences. The paper focuses on using case-based reasoning algorithms to guide students on the basis of their academic performance.

**1.3 SCOPE AND LIMITATION**

This work is intended to be a high fidelity prototype of a guidance and counseling expert system capable of suggesting courses of study to tertiary institution aspirants based on their performances in selected key subjects. This prototype was developed as a desktop application using MATLAB.

Moreover the data set used in the system contains only 1000 cases and 8 features.

**CHAPTER TWO**

**2.0 LITERATURE REVIEW**

The application of the case based reasoning in a diversity of problem scenarios has been the focus of various research works in the last few decades, some of which include the works of Maha N, Anum A, Unsa T, Jannat F.S, Rabia A, and Maria T (2014), The work of Adebayo Kolawole John, Adekoya Adewale M.andEkwonna Chinnasa (2014) and the work of M.Akhil jabbar,B.L Deekshatulu,and Priti Chandra (2013).

**2.1**

The work of Maha *et al* (2014) of the Department of Computer Sciences Kinnaird College for Women Lahore, Pakistan focused on the design of Automated Career Counselling System for Students using CBR and J48 using WEKA as a toolbox. The aim was to present an automated system that mimics a one-to-one meeting with a professional career counsellor. The system supports people in developing their own career choosing competences. The system proposes a suggestion to students regarding majors most appropriate for them. It is a formulated technique for analyzing an individual’s abilities through his/her interests and hobbies. System is designed in a way that it takes inputs from the user, matches it with the training data and yields an output. Following are the fields that the user fills as inputs:

1. Name (String)

2. Gender (Char-->F/M)

3. High School Grade (Char--->A-F)

4. Hobbies (Radio Buttons)

5. Skills (Radio Buttons)

6. IQ Grade (Char--->A-F)

The values of each field; Name, Gender, Grade in A levels/intermediate, Hobbies, Skills, IQ Grade and Current Major are utilized which shape the training set. The training set was in the form of a txt file. About 200 cases which kept increasing due to the revise logic of the algorithm CBR. A separate unit case is formed from the user’s entries in the system’s interface, which asks the user to provide input for the same fields as stated before excluding Major, and CBR is applied. Each row of training set is compared through CBR’s logic where every row of training set is compared with unit case. As each column is matched, a similarity number (3) decrements. Ultimately, the “Major” column of the row with the lowest similarity number is displayed. The Hamming distance is then computed following the algorithm;

Hamming Distance (count Matches: No. of attributes, sim: Similarity number, a: Training set’s attributes, b: given case’s attributes) 1) Check if sim=3, then, hamming distance is calculated.

2) Assign a value to count Matches according to the distance and number of attributes.

3) If the attribute of a matches with b count Match is decremented. This is done till 3rd attribute of both a and b as sim is 3.

4) If the first three attributes of ado not matches with b, then last three attributes are checked and count Matches is decremented each time it matches. 5) The difference is then calculated by dividing count Matches with 7.

6) And the major is allotted.

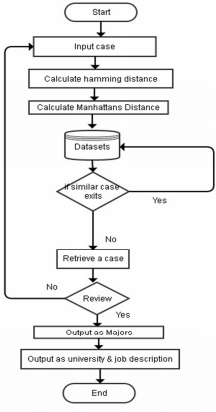
Similarly the manhattans distance is then computed thus; Manhattan Distance (sim: Similarity number, a: training set’s attributes, b: given case’s attributes)

1) Else if sim = 2, then, Manhattan distance is calculated.

2) Add up all the absolute differences of the attributes of a and b.

3) Adding the value to difference[i].

4) Update the value of difference[i] by difference[i]\*6+firstAttDist/7.



**Fig2.1 Flow Chart Showing The Flow Of The System**

The input data was selected with a keen realization of its effects on the precision of results. They colected these data from Graduate students of 5 different Universities of Lahore, Pakistan in the form of a questionnaire whence 20 majors' list was acquired to work on. It included a set of questions that recognize the personal, educational and intelligence attributes of students of particular majors. 70% of those questions determined the IQ, and the rest were divided into questions that calculate the personality and professional concerns of the pupils. The questionnaire provided with a list of hobbies that generally interest that age group, including activities resembling outdoor sports, collecting items, playing musical instruments or listening to music, creativity based activities such as reading or writing. The students were expected to choose as many hobbies that appealed them, thus determining the attribute for Hobbies. The questionnaire also included a list of skills that are generally found in professionals working in the fields to better accommodate the user to a specific career. Grade in A levels/intermediate required the target population to fill in their academic grade in FSc./FA/ICS./I.Com to yet again better understand the academic improvements of the students studying a certain major. IQ Quiz Grade is the grade that these students attained in the quiz provided along with the questionnaire to aptly judge their intelligence and IQ tendencies to undertake a major. The quiz analyzed their command on language, mathematical knowledge, observation and problem solving techniques. In Current Major, the students provided their major which aids the system to classify the similar attributes of

each major and advise careers to the users accordingly. Majors proposed may be more than one depending on the user’s input. The use case is then entered in the training set as a revised data point. After the majors are proposed, respected Universities are also proposed to the student as well as the jobs description and nature of field.

They also implemented the sysystem using Decision Tree J 48 In this system, Decision Tree J-48 is employed through Weka, a software collection of machine learning algorithms. The J48 classifier is a simple C4.5 decision tree for classification. It creates a binary tree. With this technique, a tree is constructed to model the classification process for proposed majors . The datasets collected were input in a .txt file and then converted to arff file for Weka compatibility. Once the file is loaded in Weka and J-48 is run on it, a tree is built. This tree is applied to each tuple in the data set and results in classification for that tuple. The algorithm used for constructing Decision Tree J-48, given a data set follows; Make-Decision-tree (DS: data set) 1. Classify the classes out of data set. (e.g.: A may be one class)

2. If (all elements of data set DS creates a class A)

3. Make a leaf node and label it as A 4. Else 5. Construct sub-trees out of DS 6. Repeat steps 1-5 until classification gets completed a tree view is generated after running the algorithm,with squared brackets showing the splitting criteria. This is the attribute name on which the parent node was split and the value (numeric) and nominal value (set) that has led to this child. The class value (Major in this case) in single quotes states the majority class in this node. The value in round brackets states (x of y) where x is the quantity of the majority class and y is the total count of examples in this node. The first node in the figure shows majors-taken percentage according to grade and then in this grade, according to gender.

CBR generates different outputs by giving priority to the similarities between the test case and the collected data sets. In the light of the data sets, if the test case matches exactly 6 attributes similar to the Computer major’s case, matches 5 attributes similar to the Mathematics major’s case and 4 attributes with Physics major’s case, then the system gives different priority on the basis of outputs. In this work the output is Computer Science, Mathematics and then Physics. Moreover, the revised concept led to the same result again and again showing the high accuracy of the system that is approximately 80%. While comparatively using J48, the output is a definite positive decision (yes) or negative decision (no) for a student to whether go with the result or not. Only a single output is provided and accuracy of the algorithm is 50-60%. In conclusion they determined that comparison based algorithms present a solution for this problem since they compare certain attributes of one case (student in this sense) with the previous cases collected (a student database). CBR and Decision Tree J-48 can perfectly illuminate the way for students to select a career that exactly matches their skills and IQ tendency. The results indicate that the system is capable of correctly proposing majors with approximately 80% accuracy when presented with sufficient data and features. Out of the two algorithms tested, CBR gave the highest accuracy and Decision Tree J-48 gave the lowest accuracy, and recommended that the knowledge base of the system be expanded in order to improve accuracy.

**2.2**

The work of M.Akhil et al (2013) involved the classification of heart disease using K- Nearest Neighbor and genetic algorithm. Their proposed approach combined KNN and genetic algorithm to improve the classification accuracy of heart disease data set. They used genetic search as a goodness measure to prune redundant and irrelevant attributes, and to rank the attributes which contribute more towards classification. Least ranked attributes are removed, and classification algorithm is built based on evaluated attributes. This classifier is trained to classify heart disease data set as either healthy or sick. This algorithm consists of two parts. 1) First part deals with evaluating attributes using genetic search

2) Part two deals with building classifier and measuring accuracy of the classifier. This algorithm can be stated thus;

HeartDisease()

Step1) load the dataset Step 2) Apply genetic search on the data set Step 3) attributes are ranked based on their value Step4) select the subset of higher ranked attributes Step 5) Apply (KNN+GA) on the subset of attributes that maximizes classification accuracy

Step 6) calculate accuracy of the classifier, which measures the ability of the classifier to correctly classify unknown sample.

Step 1 to 4comes under part 1 which deals with attributes and their ranking. Step 5 is used to build the classifier and step 6 records the accuracy of the classifier. Accuracy of the classifier is computed as Accuracy = no. Of samples correctly classified in test data /Total no.of samples in the test data.

The performance of this approach was tested with 6 medical data sets and 1 non medical data set. Out of 7 data sets, 6 data sets were chosen from UCI Repository and heart disease A.P was taken from various corporate hospitals in Andhra Pradesh, and attributes are selected based on opinion from expert doctor’s advice. Information about these attributes is listed in table 1.As a way to validate the proposed method, we have tested with emphasis on heart disease on A.P besides other machine learning data sets taken from UCI repository. The results acquired by KNN and GA reveals that by integrating GA with KNN will improve the classification accuracy for many data sets and especially heart disease for A.P. This prediction model helps the doctors in efficient heart disease diagnosis process with fewer attributes. Heart disease is the most common contributor of mortality in India and in Andhra Pradesh. Identification of major risk factors and developing decision support system, and effective control measures and health education programs will decline in the heart disease mortality.

**2.3**

The work of Adebayo Kolawole John,Adekoya Adewale M.andEkwonna Chinnasa(2014) was aimed at “Temperament and Mood Detection Using CaseBased Reasoning” . This implies the design of an expert system adapted to the field of psychology to help psychologists solve part of the problems in their complex domain. The system detects temperament and moods of individuals. To the intent of helping individuals who are out of reach of a professional psychologist to manage their personality and moods . The Temperament and Mood Detection System (TAMDS), developed in JAVA and employs Case-Based Reasoning technique. They acquired knowledge through different sources. They acquired cases from experts in form of their past interactions with clients and the classifications of temperaments given to those clients, this was done in order to get cases as close to the real world as possible. Also, some information was derived from interviews as well as several temperamental and psychological books. These are classified under these categories;

* The characteristics of the four basic temperaments
* The possibilities of an individual possessing more than one temperament type
* Mood disorders and symptoms
* The relationship between moods and temperament

The facts elicited as knowledge above are listed as a form of features encoded as the cases covering a broad personality trait which the system can reuse for learning and accurate prediction of users traits and moods based on some particular similarity ratio.The case representation was patterned after Schank‘s Dynamic Memory Model (DMM). Here, the case memory is a hierarchical structure of episodic memory organization packets' (EMOPs ). Specific cases which share similar attributes are placed under the same class referred to as a generalized episode (GE) which contains 3 items; Norms, Cases and Indices, where Norms are features common to all cases indexed under a GE, Indices are features which discriminate between a GE's cases and an index points to a more specific GE or directly to a case. Also, an index is composed of two terms i.e. an index name and an index value. The main task of a GE is to provide a generalized indexing structure for matching and retrieval of cases. In the case-base of TAMDS, data structures hold and access these cases. However, these data structures are a flat list of pair-represented cases. The organization of cases in TAMDS is done in such a way that it is represented as a sequential array of related characteristic traits of feelings. The process of knowledge representation in their work follows semantic model. Semantic analysis is a method for eliciting and representing knowledge about an area of interest. In this model, cases are referred as exemplars. The knowledge is represented in terms of categories, cases and index pointers, all of which form a semantic network. In the model, each case is associated with a category. An index may point to a case or a category. Cases are indexed by their attributes and are retrieved by matching these attributes to the new input problems. For each temperament type or mood state, there are different attributes. Matching cases in TAMDS starts with the first input attribute, the system scans for keywords in the input problem and retrieves cases based on these keywords. In this way, solutions are provided based on matched cases. In our work, we have employed words, descriptions, and classifications to index cases so as to enable accurate prediction of solutions.

TAMDS performs case retrieval as follows, first, a user enters a new problem case to the system in order to either check his temperament type or mood diagnosis and advice. The system activates a parser that scans the entered case and breaks it down into word groups, these word groupd are made up of keywords that really defines a temperament type as earlier given in the section under ltemperaments , non-keywords terms that forms parts of the sentences are disposed off temporarily. The selected keywords are grouped as features terms which provides a way of finding morphological variants of searched terms and can be used to match the cases in the database. The selected keywords are used to match all the features which describes earlier cases housed in the case base using a bi-directional search, to get the most relevant case from the list of cases, each hit keyword is ranked and associated with a weight computed as a function of its frequency of occurrence in the case base. Similarity matching plays an important role in Case-based reasoning systems. Different matching algorithm or measurements approaches can be applied to calculate the similarity between the feature values of a current case and an old case. In tis system, to perform similarity measurement; they employed the Nearest Neighbor algorithm . This mechanism locates the case that is most similar to the input problem. The intuition can be described as below: For input: a set of n points in d dimensions

K = (k1 k2 ……………. kn)

Desired output For each point k ϵ K, the nearest point to k is calculated using the formula (i) that follows

Similarity (C, S) (i) Where C is a current case, S is a stored case in the case base, w is the normalized weight, n is the number of attributes or features in each case, f is the index for an individual attribute/feature and sim( is the local similarity function. The usefulness of a case is estimated based on the presence or absence of certain features. These features specify which case is meant to be retrieved. In the representational approach, the cases are pre-structured and retrieved by traversing the index structure, e.g. memory organization packets (MOPS) . Similarity is assessed based on the location of the case in the indexing structure; neighbors are assumed to be similar. The nearest-neighbor method gives every feature a weight and results in a weighted sum to measure the similarity between two cases. The interface opens and invokes the knowledge acquisition phase where a user is allowed to converse with the system and enters a new case based on his/her conversation with the system. The user selects whether it is temperament or mood swing his/she wants to evaluate. While the user converse with the system, the system picks the keywords and evaluates it with the cases in the database. The system finally evaluates these attributes and selects a most similar case from past cases in the database based on Nearest Neighbor metrics adopted. The system then checks the solution of the retrieved similar case and reuses the solution to inform the user of his/her temperament or mood and can then advise the user on such. In the case of mood inquisition, the system reveals the user‘s state of mind as well as tells the user whether the mood should be maintained, managed or an expert‘s help is further needed.After the solutions have been proffered, the system checks the case against the cases in the database as to know whether a completely similar case existed or whether to save up the new case. Solutions are also saved up in the database. These solutions are called when a new user inputs a similar problem; otherwise a new solution is proffered. Therefore, the solutions are stored in the database for future access references. It functions to maintain the processes of retrieve, reuse, revise and retain of cases and user‘s information.

**CHAPTER THREE**

**Research Methodology**

**3.0 Materials and Methods**

The development of this system is designed to follow an evolutionary prototyping software process model which is described by the figure below.

Fig 3.1 The Expert System Development Process Model

Each cycle terminates at the production of a deliverable software increment, the first increment produces a software with the core features of the software system and in every cycle more features are to be added. The features include a more comprehensive case-base, detailed explanation of the recommendation, customisable interface, etc.

**3.1 Requirement Gathering**

The major tasks at this stage of development include data collection and pre-processing.

**3.1.1 Data Collection**

Here an excel dataset in .xls format containing one thousand (1000) different cases with ten(10) features representing the performance of SSS1 students in a cross section of important selected subjects offered in senior secondary schools in Nigeria was prepared. For each case the output contains a human counselor's career recommendation based on the students’ performances in the ten subjects.

**3.1.2 Data pre-processing**

The data set was normalized using the min-max normalization technique represented by the formula; , where

This normalization produced values between 0 and 1.

Normalization is a scaling or mapping technique used at pre processing stage. Where a new range is found from an existing one. When it is necessary to maintain the large variation of prediction and forecasting the Normalization technique is required to make them closer. Some of the widely used normalisation techniques include Min-Max, Z-score & Decimal scaling.

Min-Max Normalization is used here because the technique provides linear transformation on original range of data and also keeps the relationship among original data. Min-Max normalization is also relatively simple technique, and can specifically fit the data in a pre-defined boundary with a pre-defined boundary

**3.2 Planning**

There are various approaches to the design and implementation of expert systems, at this stage of development the core task was to choose an algorithm which would be both efficient and straight forward to implement. The various expert system paradigms include;

**3.2.1 Rule-Based expert systems**; A rule is a formal way of specifying a directive, recommendation, or advice. A rule is expressed as ; IF premise THEN conclusion or IF condition THEN action. The rule-based expert system is made up of; The knowledge base which contains the domain knowledge useful for problem solving. the knowledge is represented as a set of rules. When the condition part of a rule is satisfied, the rule is said to *fire* and the action part is executed. The database includes a set of facts used to match against the IF (condition) parts of rules stored in the knowledge base**.** The inference engine carries out the reasoning whereby the expert system reaches a solution. It links the rules given in the knowledge base with the facts provided in the database. The explanation facilities which enables the user to ask the expert system *how* a particular conclusion is reached and *why* a specific fact is needed. An expert system must be able to explain its reasoning and justify its advice, analysis or conclusion. The user interface is the means of communication between a user seeking a solution to the problem and an expert system.

**3.2.2Neural Networks**

This is based on pattern recognition. It is a set of interconnected relatively simple mathematical processing elements which looks for patterns in a set of examples and learns from those examples by adjusting the weights of the connections to produce output patterns. Input to output pattern associations are used to classify a new set of examples Able to recognize patterns even when the data is noisy, ambiguous, distorted, or has a lot of variation.

**3.2.3 Case-Based Reasoning**

The previous methods tried to find a compact representation of the data that can be used for future prediction. In case-based reasoning, the training examples - the cases - are stored and accessed to solve a new problem. To get a prediction for a new example, those cases that are similar, or close to, the new example are used to predict the value of the target features of the new example. Case-based reasoning can be used for classification and regression. It is also applicable when the cases are complicated, such as in legal cases, where the cases are complex legal rulings, and in planning, where the cases are previous solutions to complex problems.

The expert system approach used in this work is the case based reasoning because unlike decision trees and neural networks, relatively little work must be done offline, and virtually all of the work is performed at query time.

**3.3 Design**

This work made use of the popular case-base reasoning algorithm known as The K- Nearest Neighbour Algorithm (KNN). In KNN given some number k, given a new example, the k-training examples that have the input features closest to that example are used to predict the target value for the new example. The prediction can be the mode, average, or some interpolation between the predictions of these k-training examples, thus weighting closer examples more than distant examples. For this method to work, a distance metric is required that measures the closeness of two examples. In this work, the Euclidean distance metric is chosen from a variety of available similarity measures which include; Euclidean distance, Standardized Euclidean distance, Mahalanobis distance, City block metric, Minkowski metric, Hamming distance, Jaccard distance etc. mainly because the Euclidean distance is a two way function, i.e given a distance measure, the vector can be computed, also the Euclidean distance is easy to implement for row vectors.

Given an m-by-n data matrix X, which is treated as m (1-by-n) row vectors x, x, ..., x , the various distances between the vector xs and xt are defined as follows:

Euclidean distance:

The following chart summarises the design of the expert system in this work.

Fig 3.2 Chart Showing The Work Flow Of The system

**3.4 CODING**

This research work incorporated the use of a wide range of tools. The algorithm was implemented with MATLAB program. In the implementation the recommendations are represented using numerical values,

'ENGINEERING'= 1;

'MEDCINE/SURGERY'= 2;

'PHARMACY'= 3;

'LAW'= 4;

'ACCOUNTING'= 5;

'ARCHITECTURE'= 6;

'QUANTITY SURVEYING'= 7;

'MASS COMMUNICATION'= 8;

'TOO POOR'= 9;

'GOOD AT ALL'= 10;

Using the normalized dataset as a case-base, a GUI- based MATLAB application is developed to receive the user’s performance in the various courses into a row vector, which is also normalized following the min-max normalization technique.

The Euclidean distance between the new vector and each in the case-base is then computed. These distances are sorted in ascending order and the 100 nearest vectors in the case-base to the new vector are selected. The output recommendation for each of the neighbors in the case is considered the vote of that neighbor. The votes of each of the 100 nearest neighbors is determined and the majority vote is outputted as the recommendation for the new case .The matlab code for this work is attached at the appendix A.

**3.5 THE FIRST DELIVERABLE**

Since the intent of this research work isn’t commercial and consequently large scaled but rather showing the way in which the career counseling problem can be solved using artificial intelligence, the software developed here is a high fidelity prototype which shows the core features of the system and its machine learning capabilities, subsequent developmental cycles may incorporate other utilities such as student account/profile, various grading systems, personality analytic features etc.