

Fuzzy K-Nearest Neighbour Model for Choice of Career Path for Upper Basic School Students

Awoyelu I. O.

Department of Computer Science and Engineering, Obafemi Awolowo University, Ile-Ife, Nigeria
E-mail: iawoyelu@oauife.edu.ng

Oguntoyinbo E. O. and Awoyelu T. M.

Department of Computer Science and Engineering, Obafemi Awolowo University, Ile-Ife, Nigeri
Email: {emmanuelboyede82, awoyelu.tolu}@gmail.com

Received:04 May 2020; Accepted: 13 May 2020; Published: 08 August 2020

Abstract: Many students are faced with the challenge of deciding on a suitable career path. This is because decisions are characterized by a number of subjective judgments. Therefore, choosing a particular career path without first determining the suitability of a student, as a fundamental step, will yield an undesirable outcome. This paper aims at developing a career path decision making model for senior secondary schools. The concept of fuzzy logic was used in developing the model. Crisp sets are converted to fuzzy sets using fuzzy K- nearest neighbor algorithm method. The model was implemented in the MATLAB environment. The performance of the model was evaluated using specificity and accuracy as performance metrics. The results obtained showed the model has accuracy value of 90.22%. This result shows that the model is approximately 90% accurate. Also, it has a specificity value of 96.97%. These results show that the model provides a good support for decision making while eliminating the challenges of indecision and floundering that are characterized with choosing a career path among upper basic school students, that is, Junior Secondary School students. The model will also serve as a tool in enhancing the work of career experts.

Index Terms: Career path, decision support model, fuzzy logic, fuzzy K-nearest neighbor.

1. Introduction

Career can be defined as the occupational sequences of positions and jobs held during the lifetime of a particular person [1]. Selecting the appropriate career is one of the very important decisions that an individual must make [2]. The choice of career of an individual can satisfy his/her immediate interests, values and needs. The quality of an individual's life may also be influenced by the choice of career.

Students often face the problem of decision making in career selection all around the world [3,4]. Students who receive help to be able to explore career opportunities and planning programs of study, that relate to their career interests, are more likely to see how meaningful schools are [5]. Counselors are insufficient to handle students' request in choosing the right career [6]. [7] highlighted the problems of unawareness and unpopularity of the counselling department in different schools. The availability of a few number of human counsellors that attend to students in schools and the unavailability of counsellors in a good number of schools have further compounded the problem faced in the choice of career. As a result, majority of students selected the path of their career without proper guidance from professionals and their decisions are being motivated by the preferences of their parents and friends in choosing a career.

[8] reported that many youths go into unsuitable careers due to ignorance, inexperience and peer pressure. When this occurs, these youths constitute nuisance to themselves and their employers. They are also unable to contribute meaningfully to the society. The wrong choice of career by an individual, whether induced or by any form of false consciousness, will cause frustration in the work place. The individual may become the workman who quarrels with his tools [9]. This could lead to the loss of employment and unemployment with all its attendant consequences. These include a sense of uselessness or worthlessness, a feeling of rejection and grief [10]. [11] emphasized that when the right choice of a career is not made, the outcomes are job dissatisfaction, failure to make a success in a career and the desire to change job. As a result, making the right decision on what career to follow in life is very important for all students.

Research tends to support the fact that crystallization of career choice usually takes place in the secondary school and not in the elementary school [12]. The tertiary level of education is the level of education at which students limit themselves to a specific area of proficiency based on the foundational level of grooming while in the secondary school [13] This shows that the educational level at which career choice should be made is the secondary school. At secondary school level, selecting the right career begins during the transition from the upper basic school, that is, Junior Secondary School to the Senior Secondary School. At this stage, students will need to decide which career paths to follow.

Career paths, in the context of this study, are the different departments that are available to the upper basic school students that are transiting to the senior secondary school in order to achieve a career. A suitable career path will lead to a successful career. A successful completion of the upper basic school is a prerequisite for the second phase - the Senior Secondary School (SSS). The various career paths in the context of this study are the science, art and commercial departments.

The problems in choosing a suitable career can be categorized into three. Firstly, it leads to the possibility of an individual transferring the decision-making responsibility to others and refraining from making decision himself. Secondly, it leads to failure to achieve the best choice of career because of decision making delay. Thirdly, it leads to temporary unemployment [14]. The best career choices are not made by students because most decision made are influenced by their parents and peers' love for a particular career [15]. Other additional influencing factors are: students' perceptions about whether a subject is difficult or easy, like or dislike for subject teachers and gender. Most parents do have a definite career choice for their kids; not taking note the child's capabilities, interest and intelligence. These choices made from parental pressure are sometimes wrong and have diverted these children to psychological issues which have destroyed their future [16]. Decision on an appropriate career path for students transiting from upper basic school to senior secondary school is influenced by a number of judgements which are subjective. This leads to towing the wrong career path and/or underperformance in a chosen career. In addition to the foregoing, there are always insufficient career counsellors. There is the need to correct this precarious situation, hence this paper aims at formulating a decision support model for career path for upper basic school students.

Most published research works on career decision making are to assist senior secondary school students in choosing the right discipline as they enter into tertiary institutions. There is also dearth of literature on career path decision support systems for students. Hence, this study focuses on using students' information, Basic Education Certificate Examination (BECE) results, personality profile and past academic records to classify students into one of career path options of "Art", "Commercial", and "Science". The objective of this paper is to propose Fuzzy K-nearest Neighbour model for the classification of Career paths for Junior Secondary School students.

2. Related works

There are some published works in predicting career for students. Machine learning and data mining approaches are not new to researches in in this area.

[5] developed a decision support system (DSS) to assist counselors in identifying the right career for students after counselors would have considered the best area of specialization that fits a student. The system was implemented on Visual Basic version 6.0 and tagged Career Master. It contains four databases which are course, subject, pass and study tables. The course database contains the suitable courses that a student can study. The subject database contains all science, commercial and arts subjects that are taken by the secondary school students. The pass database contains the administrators' password. The study database contains all the questions that were used in getting the intelligent quotient (IQ) of each student and it also consists of answers to the IQ questions. An appraisal of the implementation of this system was done by making comparison with existing methods of counselling students on career path. The conclusion was based on some basic features that human counsellors use to evaluate like friends, parent's influences, IQ and so on. Past academic records were not considered in determining suitability of students for a course in the university. IQ test was based on mathematics and literature test only. The system assisted several students in identifying the right course to study before entering tertiary institutions. The language of implementation-VB is limited in terms of cross-platform compatibility. This implies that the DSS cannot be easily transferred and use in computer systems that use operating systems other than Microsoft operating system. [17] developed a DSS that used collaborative filtering technique for academic orientation. The system used dataset of students from different Spanish secondary schools. It offers guidance to counselors so as to assist their decision making in helping students select the best career path in their academic field. The system, based on the concept of collaborative recommender systems, supports academic orientation that analyzes the skills, attitudes and preferences of students. As a result, it computes relevant information that support users' decision concerning their academic future. It uses evaluative tools such as tests, essays, tasks, exercises, and so on to evaluate students. Factors such as past academic records were not considered. Past academic records are one of the criterion used in predicting career path of students in the proposed model.

[18] also proposed a DSS that guides new students to select a course in Gomal University. The system was rule-based and assisted students in selecting the most suitable faculty based on the students' capabilities. It worked like an expert system, which consists of an inference engine and a general rule base. The inference engine extracts several rules from the general rule base by making use of the C Language Integrated Production System (CLIPS) language. Rules for the DSS came from the knowledge of the Gomal university academic experts. The requirements for admission into Gomal University, the previous information about existing students and their possible results in the future also form part of the rules used by the DSS. The important criteria, capabilities and abilities in the faculties in Gomal University were collected and this information was converted into rules and facts using the CLIPS language and it was stored in the DSS knowledge base. The system was implemented using Microsoft VB because CLIPS language does not support graphical

interfaces. The system, designed using Microsoft Visual Basic, was used to measure and test the capabilities of the students like intelligence level, understanding, students past academic record, mathematical concepts, comprehension and users result. CLIPS usage allowed forward chaining and rule-based which is based on an algorithm called Rete for pattern matching [19]. The recommendation was based on the abilities module test result, background information of the students, abilities module tests result and current academic record. The DSS was able to select the best course or study or faculty for each student based on their individual capabilities and abilities that were retrieved from the test module result. The system was only useful to guide new students in the selection of a university faculty. The system also gives a general test result for determining the best major for students without considering their past academic records.

[20] developed a web-based DSS that assists end-users to be informed about studies in higher education so as to get users choose their appropriate career. Object-oriented approach with Unified Modeling Language (UML) was adopted to model the system. The system was implemented using *Hypertext Preprocessor* (PHP), MySQL and Javascript. Furthermore, Ajax technology was used to enhance the exchange of data with the database especially on-the-fly. To provide proper visualization, Google maps API was introduced into the system. Only job requirements and their relationship to higher educational studies were considered. The study was relevant in that it only gives information about higher education studies; but several factors such as students' academic records and interest that determines a student's suitability for a university course were not considered. Students' academic records are important to determine appropriate career of a student. [21] worked on university-course admission choice using fuzzy logic. This study developed a DSS for students seeking admission into the university. These students have been identified as having problem with their choice of admission. Fuzzy logic approach was adopted to linguistically and intuitively model the human reasoning and communication in the system. The fuzzy inference system of the whole process of decision was implemented. Parameterization of the model was handled using information gotten from the admission system of Nigerian University. A defuzzification module, a fuzzifier module, a knowledge base and fuzzy inference engine were the elements of the fuzzy inference DSS. A student interest and ability were incorporated into a two state variable model that was adopted. The fuzzy If-Then rules, fuzzy logical operations and membership functions were used in the fuzzy inference process. The fuzzy rules were set based on the choice viability and output variable which were extracted from the four input variables which are Choice Preference (CP), Ability (A), University Competiveness (UC) and Course Competiveness (CC). The fuzzy rules modeled how a real life academic counsellor advises a student on the best career path based on the student's ability and the possible competition in his/her area of interest. The system cannot be used in determining career path for upper basic school students because the choice of career path had already been made. This was narrowed down to specific area of proficiency in terms of university course of choice. Past academic records were not also considered by the system.

Limitations of existing works

Most published research works on career decision making are to assist senior secondary school students in choosing the right discipline as they enter into tertiary institutions. In these systems, the choice of university courses had already been narrowed down to specific areas of proficiency based on the career path already chosen. This implies that a student that chooses art career path will be limited only to courses in the art field irrespective of whether such a student is interested in science related fields. Therefore, these systems cannot assist students in upper basic school transiting to senior secondary school. There is also dearth of literature on career path decision support systems for students transiting from upper basic school to the senior secondary school.

3. Methodology

This section deals with description of the model, the dataset and the methodology adopted in this study.

A. The Proposed Model and the Dataset

The system architecture of the fuzzy based career path decision support model is presented in Fig. 1. The dataset used is the Basic Education Certificate Examination (BECE) result, past academic records and personality profiles of students of Basorun High School, Ibadan, Oyo State, Nigeria. It contains a total of ninety-two (92) students made up of 58 girls and 34 boys.

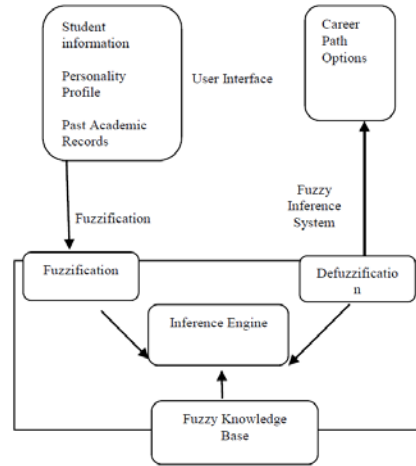


Fig. 1. System Architecture for Career Path Decision Support Model

The dataset has nine (9) attributes based on personality profile, past academic record and the seven (7) core subjects that are taken at BECE. A sample of the dataset is as shown in Table 1. The dataset's output is a score on the four categories of student classes. This study attempts to identify the rules that distinguish the four different classes, that is, 'Science', 'Art', 'Commercial' and 'Repeat'. The method for assigning classes to students is dependent on supervised machine learning and model based diagnosis. The classification is based on the identification of patterns in the data. The task is to compare observations of the system modeled with that generated by the model. If the two deviate from each other an anomaly is considered detected. Table 2 gives a sample dataset for personality profile based on interest. Fuzzy K-Nearest Neighbour (FKNN) algorithm was employed as the classifier to assist in deciding the appropriate career path for each student based on the three identified influencing factors. The training and testing parameters and conditions are defined in Table 3. Data was segmented into 4 distinct classes, Classes 1 to 4 were assigned for 'Science', 'Arts', 'Commercial' and 'Repeat', and defined in Table 4. The model estimation was carried out in the MATLAB environment by comparing the true labels with the classifier assigned crisp output values. The data was randomly stratified and the FKNN-based model was trained on the training sets to tune the parameters via stratified 2-fold cross validation in order to obtain the optimal parameter pair (k, m), and then the resulting optimal FKNN model was employed in carrying out the classification tasks.

B. Development of the Fuzzy-based Career Path Decision Support Model

Fuzzification is the first step to apply in fuzzy inference system. Most variables existing in the real world are crisp or classical variables. There is need to convert the crisp variables (both input and output) to fuzzy variables, and then apply fuzzy inference to process the data to obtain the desired output. Finally, in most cases, those fuzzy outputs need to be converted back to crisp variables to complete the desired control objectives. The basic architecture is shown in Fig. 2. This basic architecture consists of four main parts: Fuzzifier, Fuzzy knowledge Base, Inference Engine and Defuzzifier. The fuzzifier converts input variables into fuzzy values using a membership function. The inference engine is the collection of stored IF-THEN rules. This evaluates the different rules in the knowledge-base to obtain a fuzzy output. The defuzzifier then translates this output into crisp value. The function which maps fuzzy linguistic Mamdani fuzzy inference system was used to develop the fuzzy rule-based approach. Fig. 3 shows the basic Education Certificate Examination (BECE) result combined with past academic record and personality profile as the input variables to give appropriate career path options as the output variable. It consists of four operating mechanisms namely: fuzzification, fuzzy knowledge base, inference engine and defuzzification.

Fuzzification: Variables fuzzification consists of converting the clean values that were initially assigned to the input variables into their respective fuzzy logic linguistic variables. The membership function was used to specify the model mathematically. Fuzzification converts the clean value to a fuzzy input value that has a triangular membership function. This function converts the new sets to sets that are fuzzy because of their computational efficiency and formula simplification. value to an appropriate crisp value between the interval [0, 1] is called the fuzzy membership function. Using the membership function, fuzzy logic based system is able to represent and reason with a linguistic type of knowledge [22]. This process is equivalent to converting or mapping classical set to fuzzy set to varying degrees. The input is constantly a crisp numerical value that is restricted to the universe of discourse of an input variable, and the output is a fuzzy degree of membership in the qualifying linguistic set (always the interval 0 and 1). The membership function μ_A of a fuzzy set A is expressed as in Equation (1).

$$\mu_A: \chi \rightarrow [0, 1] \quad (1)$$

where χ is a universal set, and μ_A is a membership function of fuzzy set A.

Table 1. Sample Dataset for Basic Education Certificate Examination Result

REG NO	STUDENT NAME	EN	MT	BSCI	SS	CE	BT	BS
001	ABASS EZEKIEL	C	P	A	C	C	C	A
002	ABDULAHI WAR	C	P	C	P	C	C	C
003	ABIADDE TAIWO	P	P	P	P	C	P	C
005	ABIDEEN ODUNA	P	P	C	C	C	P	P
006	ABIODUN ROFIA	P	P	P	P	P	A	P
008	ABIOLA BENJAM	P	P	C	C	C	A	C
016	ADEBOYE SAMU	C	C	C	C	C	A	C
018	ADEDOKUN BOS	C	P	C	P	P	C	P
020	ADEGBOLA QOZ	C	C	C	P	C	A	C
023	ADEKUNLE AYO	C	P	P	C	C	C	P
024	ADELALU TOSIN	P	P	P	C	C	C	P
029	ADEOYE ROFIA	C	P	C	C	C	C	C
033	ADESINA DAMI	C	P	P	C	C	C	P
034	ADESINA SULIAT	C	C	C	C	C	C	P
036	ADEYEMI OREOL	P	P	P	P	P	C	P
038	ADEYEMO EMM	P	P	C	P	C	C	C
039	ADEYEMO DEBO	P	C	C	C	C	A	P
045	AFOLABI ESTHE	P	P	P	P	P	C	P
046	AFOLALU TITILA	P	C	C	C	C	A	C
050	AJIBADE TAIWO	P	P	C	C	C	C	P
052	AJIBOYE TABITH	C	C	C	C	C	A	C
054	AKANBI OYINAD	P	P	C	P	C	C	C
056	AKANDE BUKOL	C	P	C	C	C	C	C
061	AKINGBALA OPE	P	P	C	C	C	A	C

Legend: EN denotes English, MT denotes Mathematics, BSCI denotes Basic Science, SS denotes Social Studies, BT denotes Basic Technology, BS denotes Business Studies and CE denotes Civil Education.

Table 2. Sample Dataset for Personality Profile based on Interest

S/N	STUDENT NAME	AGE	SEX	OU	ME	CO	SC	PE	AR	LI	MU	SS	CL	RC
1	ABASS EZEKIEL	13	F	37	37	38	48	35	36	32	30	34	40	SC
2	ABDULAHI WAR	17	M	39	31	38	28	32	35	32	46	37	40	CO
3	ABIADDE TAIWO	15	F	29	22	25	48	32	23	29	24	41	22	SC
4	ABIDEEN ODUNA	14	F	34	26	41	41	35	38	42	40	36	43	CO
5.	ABIODUN ROFIA	15	F	31	23	25	43	39	23	29	40	47	23	CO
6.	ABIOLA BENJAM	14	F	37	31	47	30	39	25	22	23	35	44	SC
7	ADEBOYE SAMU	16	F	22	20	48	20	23	17	17	14	29	33	SC
8	ADEDOKUN BOS	13	F	34	32	45	35	40	38	41	41	38	46	CO
9	ADEGBOLA QOZ	16	F	40	22	38	24	38	26	36	25	38	24	SC
10	ADEKUNLE AYO	16	M	32	30	27	30	35	25	36	42	31	27	AR
11	ADELALU TOSIN	16	F	36	32	44	36	38	38	49	41	30	39	AR
12	ADEOYE ROFIA	14	F	24	19	26	26	37	23	29	36	16	42	CO
13	ADESINA DAMI	15	F	34	28	46	36	41	29	32	31	41	50	CO
14	ADESINA SULIAT	15	F	29	26	29	41	36	29	35	30	32	37	SC
15	ADEYEMI OREOL	16	M	26	32	35	31	36	40	43	38	33	34	AR
16	ADEYEMO EMM	14	M	34	33	36	36	37	37	31	44	38	31	AR
17	ADEYEMO DEBO	15	M	34	31	35	25	31	29	27	19	27	44	CO
18	AFOLABI ESTHE	14	F	28	30	37	42	46	32	39	27	37	46	CO
19	AFOLALU TITILA	13	M	36	35	36	35	35	39	35	33	30	38	AR
20	AJIBADE TAIWO	15	M	49	45	37	35	47	34	37	44	47	40	SC

Legend: OU- OUTDOOR; ME-MECHANICAL; CO-COMPUTATIONAL; SC-SCIENTIFIC; PE-PERSUASIVE; AR-ARTISTIC; LI-LITERARY; MU-MUSICAL; SS- SOCIAL SERVICE; CL- CLERICAL; RC-RECOMMENDED CLASS.

Table 3. Training and Testing Parameters for the Fuzzy Logic-based Classifier

	No. of Instances	No. of Features	No. of Classes
Training set	46	9	4
Test set	46	9	4

Table 4. Different Categories of Data and Descriptions

S/No.	Data Category	Description
1	Class 1	Science
2	Class 2	Art
3	Class 3	Commercial
4	Class 4	Repeat

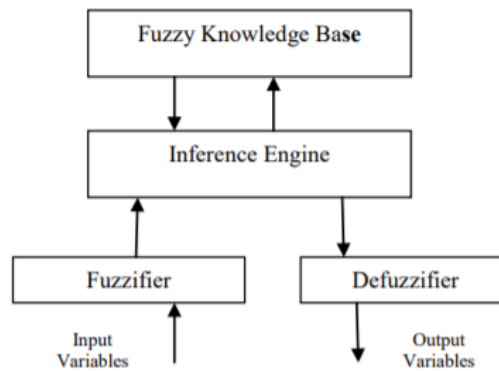


Fig. 2. A Basic Architecture of Fuzzy Logic System

Table 5 depicts the fuzzy set for the input variables - BECE showing linguistic variables which are distinction, credit, pass, and fail, its notations, ranges and crisp values for the fuzzy. These linguistic variables assist in the representation of the transition which is from a higher level to a lower level. They are words from natural language instead of numerical values.

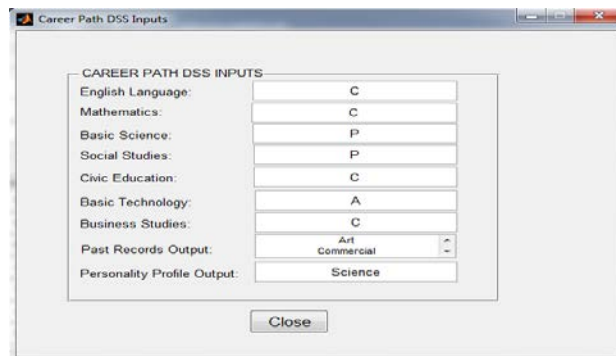


Fig. 3. Sample Input Records of a Student

Table 5. Fuzzy Set for BECE

Linguistic Variable	Notation	Range	Crisp Values
Distinction	A	70 - 100	1
Credit	C	50 -59	2
Pass	P	40 - 49	3
Fail	F	0 -39	4

Table 6 is the fuzzy set for the past academic records, showing linguistic variables which are excellent, very good, good, fair, poor, and fail, its notations and crisp values. The linguistic variables denote the transition from a higher level to a lower level. Table 7 is the fuzzy set for the input variables, students' personality profile showing the personality

type, with its corresponding career path and the linguistic variable such as dislike very much, dislike, indifferent, like, like very much. Membership grade within interval [0,1] is allowed for members in the fuzzy set.

The k-nearest neighbor (KNN) algorithm which is a machine learning algorithm, which carries out its classification tasks by assigning the class or output of a new sample by referring to a very frequent class within its k nearest neighbors.

Table 6. Fuzzy Set for Past Academic Records

Linguistic Variable	Notation	Range	Crisp Values
Excellent	A	70-100	1
Very Good	B	60-69	2
Good	C	50-59	2
Fair	D	45-49	3
Fail	E	40-44	3
	F	0-39	4

Table 7. Fuzzy set for Students' Personality Profile

Questions (Section A-J)	Personality Type	Career path	Linguistic Variable
1	Outdoor	Science	Like very much, like, indifferent, dislike, dislike, dislike very much
2	Mechanical	Science	Like very much, like, indifferent, dislike, dislike, dislike very much
3	Computational	Science	Like very much, like, indifferent, dislike, dislike, dislike very much
4	Scientific	Science	Like very much, like, indifferent, dislike, dislike, dislike very much
5	Persuasive	Art	Like very much, like, indifferent, dislike, dislike, dislike very much
6	Artistic	Art	Like very much, like, indifferent, dislike, dislike, dislike very much
7	Literary	Art	Like very much, like, indifferent, dislike, dislike, dislike very much
8	Musical	Art	Like very much, like, indifferent, dislike, dislike, dislike very much
9	Social Service	Commercial	Like very much, like, indifferent, dislike, dislike, dislike very much
10	Clerical	Commercial	Like very much, like, indifferent, dislike, dislike, dislike very much

The FKNN is a feature-based classifier system. In this study, features were extracted from the segmented training dataset (50% for training and 50% for testing) and fed into the classifier which assigns a user-defined class label to the obtained results. The FKNN classifier system possesses the capability to adjust itself to the ambiguity in the training and test data. The FKNN algorithm is as shown in Algorithm 1.

Algorithm 1: The FKNN Algorithm

Input: set of labelled training document X, such that $\{x_j | j = 1, 2, \dots, n\}$ with the testing pattern y.

Output: Class label of y and confidence for each class label.

1. For $i = 1, 2, \dots, n$
 2. Compute the distance from x_i to y using the Euclidean distance;
 3. If i to k
 4. Include x_i in the set of k nearest neighbors;
 5. Else If (x_i is closer to y than any previous nearest neighbors)
 6. Discard the farthestmost of the k nearest neighbors;
 7. Include x_i in the set of k nearest neighbors;
 8. End If
 9. End for
 10. For $c = 1$ to C
 11. Compute $u_i(x)$ using Equation 4;
 12. End For
 13. Crisp class label of y is assigned to the class with which it has the highest membership value according to Equation 5.
-

C. Fuzzy Knowledge Base

The fuzzy knowledge base consists of the rule base and the database. Fuzzy rules were created basically from the knowledge or information elicited from career experts. The rule base consists of a collection of rules which represent the knowledge about the application domain specified by the expert. The career counselor acts as the domain expert by providing data about personality profile and academic records and the rules required to determine appropriate career path. The <action> and <condition> parts of each rule are represented with the consequent antecedent parts, respectively. The antecedent part can be very simple or could be a complex logical combination of several conditions and more than one action may be specified in the consequent part. The rules are constructed and set the basis for deriving the values of the output variables. It models how a career expert will advise a student by comparing the basic education certificate examination result, past academic records and personality profile. A sample input data is as shown in Fig. The database component contains an ensemble of fuzzy sets membership functions that are used in the rules by fuzzy. The rules used by fuzzy include the following;

```
IF (((BECE ((en,'A') | (en,'C')) & ((mth,'A'))|(mth,'C')) & ( (bsci,'A') | ((bsci,'C')) & (ss,'A')) | ((ss,'C')) & ((bt,'A') | (bt,'C')) & ((ce,'A') | (ce,'C')) & ((bs,'A') | (bs,'C')) | ((pastacademicrecords ((en,'A') | (en,'C')) & ((mth,'A'))|(mth,'C')) & ( (bsci,'A') | (bsci,'C')) & ((ss,'A') | (ss,'C')) & ((bt,'A') | (bt,'C')) & ((ce,'A') | (ce,'C')) & ((bs,'A') | (bs,'C')))) THEN
  careerpath is 'Science-Art-Commercial' ELSE IF (((BECE ((en,'A') | (en,'C')) & ((mth,'A'))|(mth,'C')) & (ss,'A')) | ((ss,'C')) & ((ce,'A') | (ce,'C')) & ((bs,'A') | (bs,'C')) | ((pastacademicrecords ((en,'A') | (en,'C')) & ((mth,'A'))|(mth,'C')) & (ss,'A')) | ((ss,'C')) & ((ce,'A') | (ce,'C')) & ((bs,'A') | (bs,'C')))) THEN careerpath is 'Art-Commercial'
  ELSE IF (((BECE ((en,'A') | (en,'C')) & ((mth,'A'))|(mth,'C')) & ((ce,'A') | (ce,'C')) & ((bs,'A') | (bs,'C')) | ((pastacademicrecords ((BECE ((en,'A') | (en,'C')) & ((mth,'A'))|(mth,'C')) & ((ce,'A') | (ce,'C')) & ((bs,'A') | (bs,'C'))
  THEN careerpath is 'Commercial' ELSE IF (((BECE ((en,'A') | (en,'C')) & ((mth,'A'))|(mth,'C')) & (ss,'A')) | ((ss,'C')) & ((ce,'A') | (ce,'C')) & ((pastacademicrecords ((en,'A') | (en,'C')) & ((mth,'A'))|(mth,'C')) & (ss,'A')) | ((ss,'C')) & ((ce,'A') | (ce,'C')) THEN careerpath is 'Art' ELSE IF (((BECE ((en,'A') | (en,'C')) & ((mth,'A'))|(mth,'C')) & (bsci,'A')) | ((bsci,'C')) & ((bt,'A') | (bt,'C')) | ((pastacademicrecords ((en,'A') | (en,'C')) & ((mth,'A'))|(mth,'C')) & (bsci,'A')) | ((bsci,'C')) & ((bt,'A') | (bt,'C')) THEN careerpath is 'Science' ELSE IF (((BECE ((en,'A') | (en,'C')) & ((mth,'A'))|(mth,'C')) & ( (bsci,'A') | ((bsci,'C')) & (ss,'A')) | ((ss,'C')) & ((bt,'A') | (bt,'C')) & ((ce,'A') | (ce,'C')) | ((pastacademicrecords ((en,'A') | (en,'C')) & ((mth,'A'))|(mth,'C')) & ( (bsci,'A') | (bsci,'C')) & ((ss,'A') | (ss,'C')) & ((bt,'A') | (bt,'C')) & ((ce,'A') | (ce,'C')))) THEN careerpath is 'Science-Art' ELSE IF (((BECE ((en,'A') | (en,'C')) & ((mth,'A'))|(mth,'C')) & ( (bsci,'A') | ((bsci,'C')) & ((bt,'A') | (bt,'C')) & ((ce,'A') | (ce,'C')) | ((pastacademicrecords ((en,'A') | (en,'C')) & ((mth,'A'))|(mth,'C')) & ( (bsci,'A') | ((bsci,'C')) & ((bt,'A') | (bt,'C')) & ((ce,'A') | (ce,'C')))) THEN careerpath is 'Science-Commercial' ELSE IF (((BECE ((en,'F') & ((mth,'F')) THEN ' repeat class '
  End
```

End
LEGEND:

en- ENGLISH; mth- MATHEMATICS; bsci-BASIC SCIENCE; ss-SOCIAL STUDIES; ce-CIVIIC EDUCATION; bs-BUSINESS STUDIES; bt-BASIC TECHNOLOGY

Personality Profile Rules

```
if (pt, 'outdoor')
  pp = 'Science'
else if (pt, 'mechanical')
  pp = 'Science'
else if (pt, 'computational')
  pp = 'Science'
else if (pt, 'scientific')
  pp = 'Science'
else if (pt, 'persuasive')
  pp = 'Art'
else if (pt, 'Artistic')
  pp = 'Art'
else if (pt, 'literary')
  pp = 'Art'
else if (pt, 'musical')
  pp = 'Art'
else if (pt, 'social')
  pp = 'Commercial'
else if (pt, 'clerical')
  pp = 'Commercial'
end
```

- If (Career_Path_Options is science or art or commercial class) and (if personality_profile is science) then (Career_Path_Options is science class)

- b. If (Career_Path_Options is art or commercial class) and (if personality_profile is commercial) then (Career_Path_Options is commercial class)
- c. If (Career_Path_Options is commercial class) and (if personality_profile is commercial) then (Career_Path_Options is commercial class)
- d. If (Career_Path_Options is art class) and (if personality_profile is art) then (Career_Path_Options is art class)
- e. If (Career_Path_Options is science class) and (if personality_profile is science) then (Career_Path_Options is science)
- f. If (Career_Path_Options is science or art class) and (if personality_profile is science) then (Career_Path_Options is science)
- g. If (Career_Path_Options is science or commercial class) and (if personality_profile is science) then (Career_Path_Options is science class)
- h. If BECE (English_language is P) and (mathematics is P) then repeat class.

D. Inference Engine

The Inference engine has the rule based decision support capacity and information processing component of the career path decision support system for schools. It scans through the fuzzy knowledge base to evaluate the different rules in the knowledge base. A fuzzy response is generated from these fuzzified inputs using a rule based inference system. The inference engine is the mechanism that performs the reasoning and searching in the fuzzy knowledge base. It implements the fuzzy reasoning by combining the fuzzified inputs with the rules to obtain a fuzzy output. The resulting data collected from the fuzzy classification system is a combination of aggregate and single fuzzy set when the inference process is complete.

E. Defuzzification

When the inferencing is over, there is need to compute a single value to represent the outcome. This process is called defuzzification. It is also the conversion of an output variable to original crisp values from the fuzzy linguistic values. The result gotten for the rules are defuzzified and aggregated to get uniquely new values. In this study, the defuzzification method used is the centre of area, otherwise known as Centroid of Area (COA) defuzzification technique. Centroid of area zCOA is defined in Equation (2).

$$zCOA = \frac{\int z\mu_A(z)zdz}{\int z\mu_A(z) dz} \quad (2)$$

where $\mu_A(z)$ is the aggregated output MF. This is the most widely adopted defuzzification strategy, which is reminiscent of the calculation of expected values of probability distributions.

F. Classification Phase Using FKNN

The model implemented its classification tasks by making use of the randomly stratified data. It contains two main sub-modules. In FKNN classification, the first step is to set up all the model parameters. Because the fuzzy strength parameter has a major impact on the performance of the FKNN-based model, an experimental outline was drawn in order to select the optimal fuzzy strength parameter for the FKNN classifier. An interval of (1, 2) was chosen and continually increased with steps of 0.01 for the fuzzy strength parameter m , and subsequently classification performance was confirmed by way of the 2-fold CV analysis on various values of nearest neighbors k . For every value of m in the interval (1, 2), the average accuracy obtained by FKNN by the way of Cross Validation (CV) analysis, and eventually, the one with the highest average accuracy was selected as the optimal fuzzy strength parameter. The best strength parameter of the fuzzy is chosen, the classifier for the FKNN was used so as to calculate the accuracy of the classification and then an average of the results was taken. The pseudo-code for the classification phase is given in Algorithm 2.

Algorithm 2: The pseudo-code for the classification phase

*/*performance estimation by using n-fold CV where $n = 2$ */*

1. *Begin*
 2. *For $i = 1:M_{max}$*
 3. *For $j = 1:k$*
 4. *Training set = $k - 1$ subsets;*
 5. *Test set = remaining subset;*
 6. *Train the FKNN model on the training set to find the optimal fuzzy strength parameter m when the neighborhood size k is set from 1 through 15;*
 7. *Test it on the test set and assigns the accuracy to $V(j)$, where V is a vector whose element is the corresponding accuracy obtained by each folder;*
 8. *End for*
 9. *Compute the mean value of vector V , and store the mean CV accuracy in the vector $M(i)$;*
 10. *End for*
 11. *Get the optimal m value whose corresponding mean CV accuracy is the highest in $M(i)$;*
 12. *End*
 13. *Begin*
 14. *For $l = 1:k$*
 15. *Training set = $k - 1$ subsets;*
 16. *Test set = remaining subset;*
 17. *Train the FKNN model on the training set using the obtained optimal parameter combination;*
 18. *Test it on the test set and save the mean CV accuracy;*
 19. *End for*
 20. *Return the average classification accuracy rates of FKNN over l test set.*
 21. *End*
-

The fuzzy inference system for career path decision support system is shown in Fig. 4. It involves the fuzzification of all input variables which are personality profile, past academic records and each compulsory seven subjects in BECE. All the input variables were fuzzified using triangular membership function because of its simplicity and clarity. It also involves the development of the rule base for the inference engine, aggregation of the results processed by the inference engine for the inputs value presented to the fuzzy inference system and defuzzification of the output using triangular membership function.

Fig. 5 shows the membership function of past academic records inputs showing the four different classes for the different career path options. The membership function is a curve that allows a graphical representation of a fuzzy set. It defines how each point in the input space is mapped to a membership value between 0 and 1. A membership function is used to quantify a linguistic term. Fig. 6 and Fig. 7 show the triangular membership function of BECE subjects - Mathematics and Basic Science inputs and the corresponding range for each notation, F (Fail), C (Credit), P (Pass), and A (Distinction). Fig. 8 shows the output membership function for the career path. The rule base developed for the inference engine is shown in Fig. 9. These rules obtained from career experts are expressed as linguistic rules. The rule viewer of the career path decision support system in Figure 10, shows the inputs and defuzzified outputs. The input values can be changed by clicking and dragging the input vertical lines to give defuzzified outputs.

G. System Evaluation

In order to properly evaluate the performance of the fuzzy inference system classifier model, a stratified 2-fold cross validation model was used that allows the folds to be selected to contain same class labels proportions. The input data was partitioned into two equal datasets.

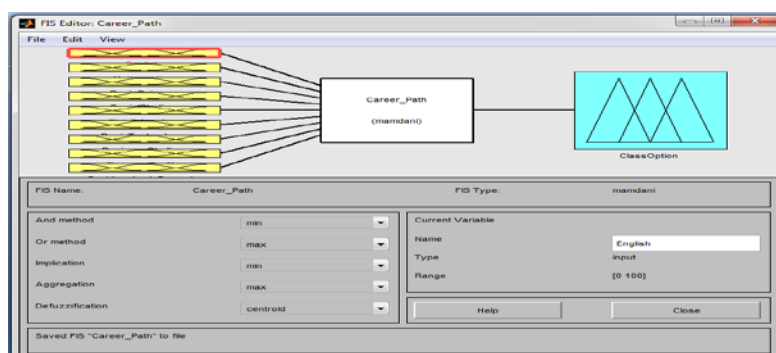


Fig. 4. Fuzzy Inference System for Career Path Decision Support System

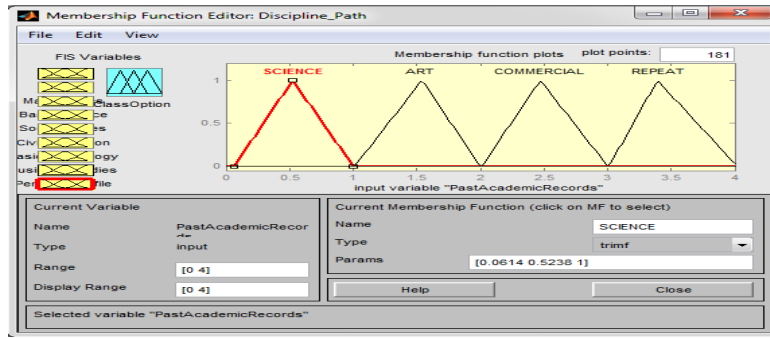


Fig. 5. Membership Function for Past Academic Records

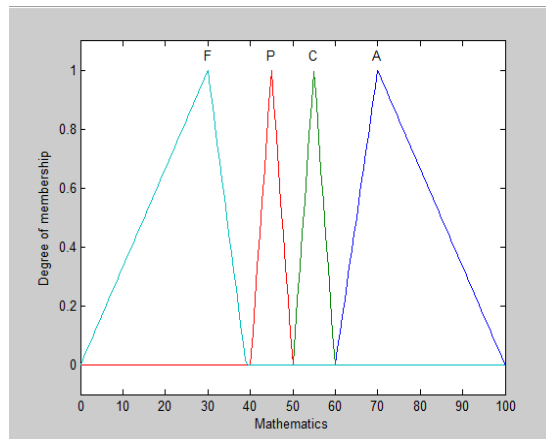


Fig. 6. Membership Function for BECE (Mathematics)

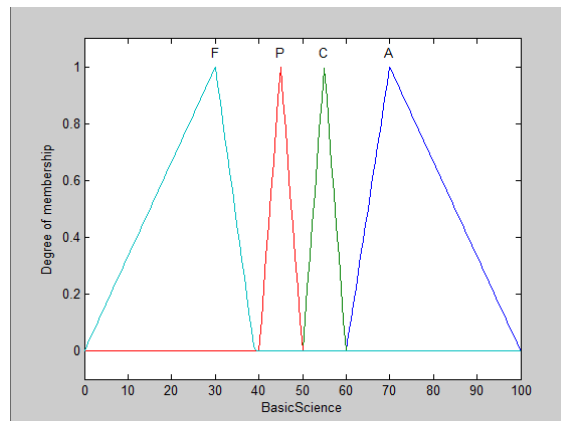


Fig. 7. Membership Function for BECE (Basic Science)

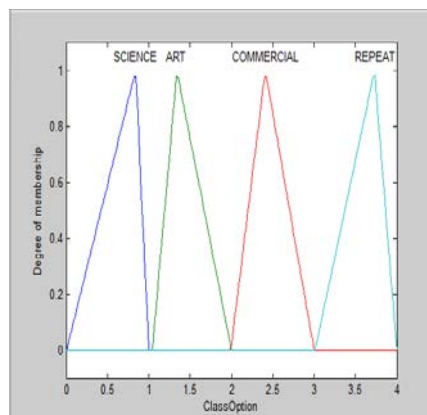


Fig. 8. Output Membership Function for Career Path

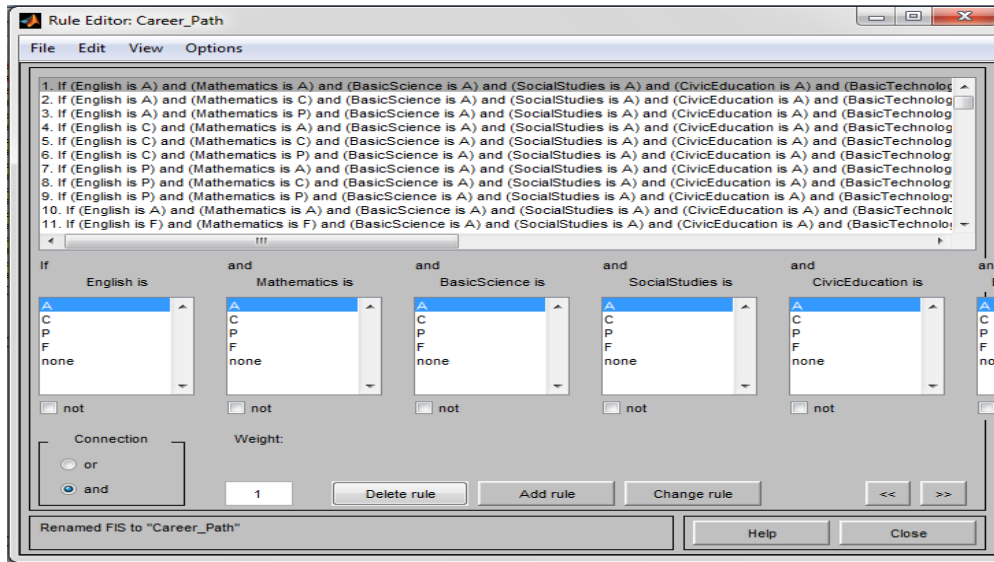


Fig. 9: Snapshot of some rules developed for the inference engine

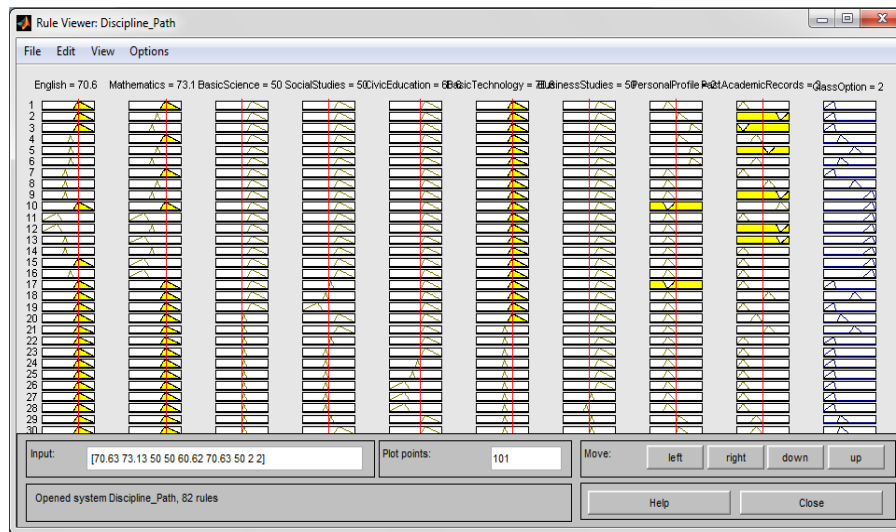


Fig. 10. Rule Viewer showing the Inputs and Defuzzified Output

Accuracy and specificity were used to test the performance of the fuzzy based career path decision support model. Accuracy is computed using the number of times a prediction is correct divided by the total number of possible outcomes. Specificity is computed as the predictions that are negatively correct divided by the total negatives. The performance of the model was evaluated using the data in the confusion matrix. This helps to report the number of false positives (FP), true negatives (TN), false negatives (FN) and true positives (TP) for a four class classifier. The rows in the confusion matrix is used to represent the actual class while the column is used to represent a predicted class. FP is the number of students whose career path was incorrectly predicted; TN is the number of students whose career path was correctly predicted; FN is the number of students whose career path was incorrectly predicted and TP is the number of students whose career path was correctly predicted. The predicted accuracy and specificity are calculated as follows:

Predicted accuracy =
$$\frac{(TP + TN)}{(TP + FN + FP + TN)} * 100\%$$
 (3)

Specificity =
$$\frac{(TN)}{(FP+TN)*100\%}$$
 (4)

The total number of cases were forty-six (46), that is the dataset used comprised of records of 46 students. A stratified 2-fold cross validation method was used, therefore the dataset was divided into two – Fold 1 and Fold 2 for cross-validation. At the first iteration, Fold 1 was used for the training set and Fold 2 as the testing case. At the second iteration, Fold 2 was used as the training set while Fold 1 was used as the testing set. The results of predicted and actual classes are as shown in the confusion matrix in Fig 11 for Fold 1 and Fig. 12 for Fold 2. The accuracy results for Fold

1 and Fold 2 are as shown in Table 8. Fold 1 has an accuracy of 91.3% while Fold 2 has an accuracy of 89.13%. The average accuracy of the model is 90.22%. This implies that the model is approximately 90% accurate in deciding appropriate career path for students. A sample output of the model is as shown in Fig. 11.

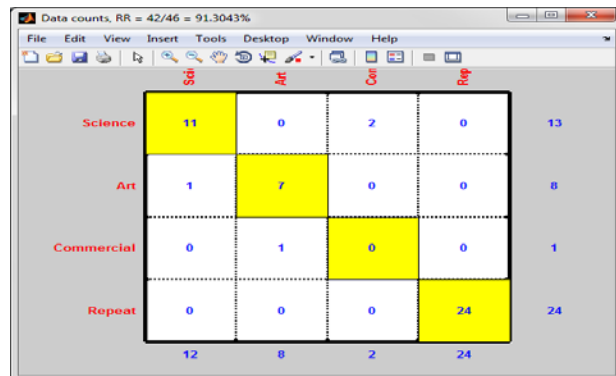


Fig. 11. Confusion Matrix displaying the Crisp Outputs for Fold 1

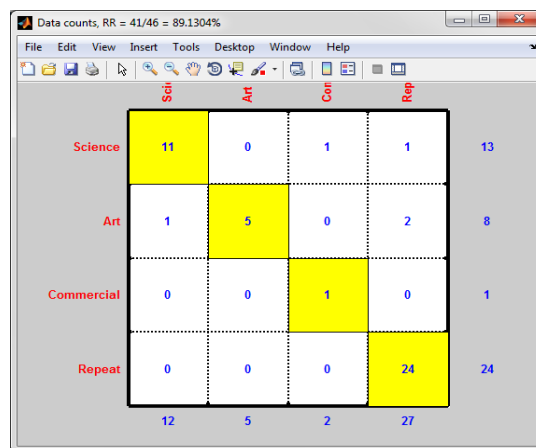


Fig. 12: Confusion Matrix displaying the Crisp Outputs for Fold 2

Table 8. Accuracy Results for Fold 1 and Fold 2

	Accuracy (%)
Fold 1	91.3
Fold 2	89.13
Average Accuracy	90.22
Specificity	96.97

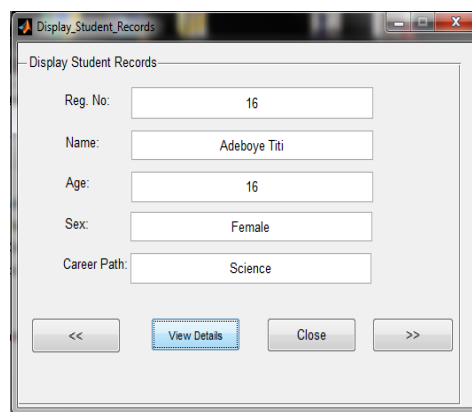


Fig. 13. A Sample Output of Career Path for a Student

4. CONCLUSION

Many students are faced with the challenges of deciding on suitable career path because decisions are characterized by a number of subjective judgements. Therefore, choosing a particular career path without first determining the suitability of the student as a fundamental step will yield an undesirable outcome. This study addressed this issue by formulating a career path decision making model. Presently, in some schools, the BECE result is the only criterion used in deciding the appropriate career path for students transiting from junior secondary school to senior secondary school while for some the decisions are influenced by parents or peers. The proposed model would be able to decide career path for students using BECE result, past academic records and personality profile. The resulting system will also serve as a tool in enhancing the work of career experts.

References

- [1] Ezeani, N. (2013). Career Choice: A Basic Issue in Primary and Secondary School Level. *Arabian Journal of Business and Management Review (Nigerian chapter)*, 1 (2): 18-28.
- [2] Gati, I. and Asher, B. (2001). Differential Diagnosis and Treatment of Career Indecision. *Personal and Guidance Journal*, 62 (1): 27-29.
- [3] Razak, T., Hashim, M., Noor, N., Hazwam, I., and Halim, A. (2014). Career Path Recommendation System for UiTM Perlis Students Using Fuzzy Logic. 5th International conference on Intelligent and Advanced Systems, pp 1-6.
- [4] Bandura, A., Barbaranelli, C., Caprara, G., and Pastorelli, C. (2001). Self-efficacy Beliefs as Shapers of Children's Aspirations and Career Trajectories. *Child Development*, 72: 187-206.
- [5] Balogun, V. and Thompson, A. (2009). Career Master: A Decision Support System for Guidance and Counseling, *The Pacific Journal of Science and Technology*, 10(2): 337-354.
- [6] Otta, F. and Njoku, O. (2012). Self-Concept and Vocational Interest among Secondary School Students (Adolescents). *Asian Journal of Social Sciences & Humanities*, 1(4): 37-48.
- [8] Abisoye, O., Alabi, I., Ganiyu, O., Abisoye, O., and Omokore, J. (2015). A Web-Based Career Guidance Information System for Pre-Tertiary Institution Students in Nigeria. *International Journal of Scientific Research in Science, Engineering and Technology (IJSRSET)*, 1(3): 229-240.
- [9] Issa, A. and Nwalo, K. (2008). Factors Affecting the Career Choice of Undergraduates in Nigerian Library and Information Science Schools. *African Journal of Library, Archives and Information Science*, 18(1): 23-31.
- [10] Ezeani, N. (2013). Career Choice: A Basic Issue in Primary and Secondary School Level. *Arabian Journal of Business and Management Review (Nigerian chapter)*, 1 (2): 18-28.
- [11] Anigbogu, M.A. (1988). *Foundation of Guidance and Counselling*. Enugu: Academic Publishing Co.
- [12] Uba, E. (1997). The Factors Affecting Career Attitude Maturity of Senior Secondary School Students in Nnewi Education Zone of Anambra State. Unpublished M.Ed. Dissertation submitted to Nnamdi Azikiwe University, Awka.
- [13] Osuala, E. (1987). *Principles and Practice of Business Education*. Uruowulu-Obosi: Pacific publishers.
- [14] Ogunlade, J. and Akeredolu, S. (2012). Influence of Counselling on Career Preference among Secondary School Students in Ekiti State, Nigeria. *Journal of Sociology, Psychology and Anthropology in Practice*, 4 (3): 26-30.
- [15] Gati, I., Saka, N., and Krausz, M. (2001). Should I Use a Computer-assisted Career Guidance System? It Depends on Where Your Career Decision-making Difficulties Lie. *British Journal of Guidance & Counselling*, 29(3): 301-321.
- [16] Kelechie, L. and Ihuoma, C. (2011). The Role of Guidance Counsellors in the Career Development of Adolescent and Young Adult with Special Needs. *British Journal of Arts and Social Sciences*, 2(1): 51-62.
- [17] Kinanee, B. (2009). Factors in the Career Decision-making of Nurses in Rivers State of Nigeria: Implications for Counseling. *Journal of Psychology and Counseling*, 1(8):134-138.
- [18] Emilio, J. and Luis, M. (2009). A Web-Decision Support System based on Collaborative Filtering for Academic Orientation: Case Study of the Spanish Secondary School. *Journal of Universal Computer Science*, 15(14): 2786-2807.
- [19] Aslam, M. and Khan, A. (2011). A Proposed Decision Support System / Expert System for Guiding Fresh Students in Selecting a Faculty in Gomal University, *Ind. Eng. Lett.* 1(4): 33-41.
- [20] Zhengxin, C. (2000). *Computational Intelligence for Decision Support*. Florida: CRC press.
- [21] Kostoglou, V., Ploskas, N., Vassilakopoulos, M., and Tsantopoulou, V. (2014). Analysis and Design of a web-based decision support system for choosing higher education studies. *Yugoslav Journal of Operations Research*, 24 (3): 399-414.
- [22] Oladokun, V. and Oyewole, D. (2015). A Fuzzy Inference based Decision Support System for Solving the University-Course Admission Choice Problem. *International Journal of Computer Applications*, 112 (3): 1-7.
- [23] Macwan, N. and Sajja, S. (2014). Fuzzy Logic: An Effective User Interface Tool for Decision Support System. *International Journal of Engineering Science and Innovative Technology (IJESIT)*, 3 (3): 278-283.

Authors' Profiles



Awoyelu I.O. is an Associate Professor of Computer Science in the Department of Computer Science and Engineering, OAU, Ile-Ife, Nigeria. She holds B.Sc, M.Sc and PhD degrees in Computer Science from the Department of Computer Science and Engineering, Obafemi Awolowo University, Ile-Ife, Nigeria. Her research interests are Data Analytics, Recommender Systems and data warehouses.



Oguntoyinbo E. O. holds B.Sc degree in Computer Science and M.Sc degree in Computer Science from the Department of Computer Science and Engineering, Obafemi Awolowo University, Ile-Ife, Nigeria.



Awoyelu T. M. holds B.Sc degree in Computer Science from the Department of Computer Science, Osun State University, Osogbo and M.Sc degree in Intelligent Systems Engineering from the Department of Computer Science and Engineering, Obafemi Awolowo University, Ile-Ife, Nigeria. She is currently a PhD student of Intelligent Systems Engineering in the Department of Computer Science and Engineering, Obafemi Awolowo University, Ile-Ife, Nigeria.

How to cite this paper: Awoyelu I.O., Oguntoyinbo E. O., Awoyelu T. M., " Fuzzy K-Nearest Neighbour Model for Choice of Career Path for Upper Basic School Students ", International Journal of Education and Management Engineering (IJEME), Vol.10, No.4, pp.18-32, 2020. DOI: 10.5815/ijeme.2020.04.03