

# International Journal of Advance Research in Computer Science and Management Studies

Research Article / Survey Paper / Case Study

Available online at: [www.ijarcsms.com](http://www.ijarcsms.com)

## *Expert system for Career Selection: A Classifier Model*

**Prof. M. L. Waghmode<sup>1</sup>**  
Bharati Vidyapeeth, IMRDA,  
Sangli, India

**Dr. P. P. Jamsandekar<sup>2</sup>**  
Bharati Vidyapeeth, IMRDA,  
Sangli, India

*Abstract: An Expert System (ES) is a computer program that simulates decision making power of human expert. ES is used in many sectors viz; medical, agriculture, mining, education, industry etc. Career plays an important role in student's life. In India the educational pattern is 10+2+3, it means students entering to career oriented courses after 10th or 12th. In India career related decisions are taken after 10th standard. Selection of proper career stream consists of step by step process. This process considers different parameters for assessing student's strength and weaknesses. In this paper researcher describes the role of expert system in career selection. Researcher collected domain knowledge from domain experts, and designed a framework of expert system for career selection. Researcher used different machine learning classifier algorithms for classifying data into different career streams using Weka tool. The performance of these classifiers is described in this paper. It is found that ID3, PART and PRISM classifier algorithms gives 100% accuracy.*

*Key words: Expert System; Domain knowledge; Machine Learning Algorithms; Weka.*

### I. INTRODUCTION

Expert systems (ES) is a computer program that represents and reasons with knowledge of a specialized domain with a view to solve problems or giving advice. ES has proven to be effective in a number of problem domains which normally requires human expertise. ES uses symbolic knowledge representation in terms of rules, networks or frames. Career is an artefact that has emerged within the broader framework of the human activity called work [1]. Career selection is one of many important choices students will make in determining future plans.

Weka stands for Waikato Environment for Knowledge Analysis. Weka is a collection of machine learning algorithms for data mining tasks. It contains tools for data pre-processing, classification, regression, clustering, association rules, visualization and forecasting. It is used for applications, research and education. [2]

### II. LITERATURE REVIEW

#### A. Career selection:

Hall (1976) defines career is the individually perceived sequence of attitudes and behaviours associated with work-related experiences and activities over the span of the person's life”.

Career is defined by the Oxford English Dictionary as a person's "course or progress through life or a distinct portion of life". Career is understood to relate to a range of aspects of an individual's life, learning and work. The term career is used to describe an occupation or a profession that usually involves special training or formal education, and is considered to be a person's lifework.[3]

For career guidance many organizations, consultancies, counsellors are guiding to students for selecting their proper career options. For this purpose they are considering different parameters for guiding students.

#### B. Career Selection process:

Career counsellors follow their own methods for giving career guidance to students. It means counsellors considers different factors for assessing student. Some uses aptitude test, psychology test, personality test, skill tests, student's background information, IQ, interests, etc., from this information they are guiding the students about career selection options.

In Indian educational system students select their career stream after their secondary school (S.S.C.), and hence it is important to make correct decision at this phase. In Maharashtra students especially those stay at backward areas are not getting proper guidance from their parents and environment, so it makes them difficult to take career decisions. And hence SSC is turning point in student's life, and for selecting proper stream they need proper guidance from teachers, counsellors, parents etc.

#### *C. Role of Expert system:*

Expert systems are a system that uses domain knowledge stored inside a computer to solve problems that require human expertise for solving. There are some limitations in solving problems with the help of human domain experts viz; less number of expert counselors, unavailability of human experts at anytime and anywhere, inaccessibility. It is very difficult to replace human expert, with time human expert knowledge may get damage, its speed and efficiency changes, it costs very high, hence expert system is useful for either to replace or help the human expert (counselors). Hence expert system enable the use of expertise at any time anywhere, to reduce operational cost by automating a routine task that requires timely human expertise. Today there are N number of choices in front of students for selecting their career path; hence one should assess himself before entering the career oriented course. Acquiring the knowledge of self-assessment and select perfect career is challenging activity. Now a days students take advice of psychologist / career counselors for career guidance. As number of counselors is less and acquiring knowledge of career selection is not affordable, hence the use of an expert system for career selection is the best option. An expert system can retain the knowledge of experts and maintain for future reuse.

#### *D. Factors affecting Career Selection:*

Though there are different factors considered in career selection, still majorly considered factors are narrated as below.

- Ability: Skills to do something.
  - It is also called aptitude tests and are generally standardized. These tests are designed to assess student's logical reasoning or thinking capabilities. They consist of a number of multiple choice questions and are strictly timed. [4]
  - These tests can be used to test variety of skills mental aptitude, problem solving, and knowledge of a particular subject, reasoning ability, and general intelligence and so on. These are used to find the suitability of a candidate for a given job role. These can be quantitative, psychological, and verbal or may take other forms as well.
- Attitude: Tendency to respond positively or negatively towards a certain idea, object, person, or situation
- Personality: It refers to the pattern of thoughts, feelings, social adjustments, and behaviours consistently exhibited over time that strongly influences one's expectations, self-perceptions, values, and attitudes.
- Interest: Interests test is used to tell more on who you are and what suits you. They provide information necessary for making the right career choice. Interest tests help you define your interests and determine what you like most. This could help you when making a career choice.

### III. OBJECTIVES

In order to design an expert system researcher has set following objectives for study.

- To design a framework of expert system for career selection.
- To collect domain knowledge and sample cases from domain expert.
- To create dataset.
- To generate model using different machine learning algorithms.
- To identify most suitable algorithm for generating rules.

### IV. FRAMEWORK FOR DESIGN OF EXPERT SYSTEM

Framework for design of expert system for career selection is shown in the following Fig. 1

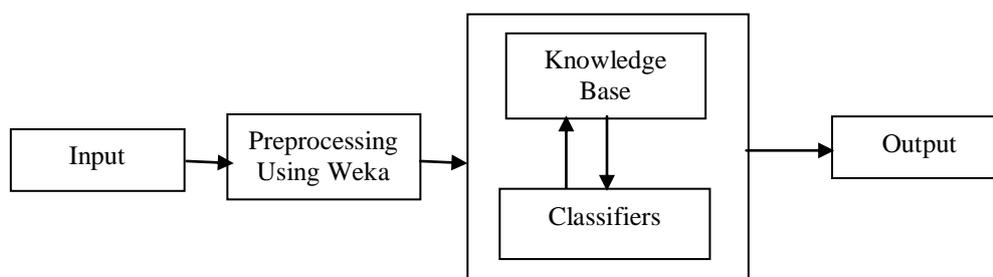


Fig. 1 Framework of Expert System for career selection

The above framework consists of different components viz; input, knowledge base, inference engine and output. Input component consists of different parameters generated from scores of different tests, these input values are then pass to the Weka tool. Here data mining preprocessing is done using Weka preprocessing filters. After getting processed data machine learning classification algorithms are used to classify into proper classes. Here different classifiers are used. These classifiers infers some rules which will be treated as knowledge base. Finally output in terms of career options are displayed as different classes viz; class a, b, c, d, e, f and g.

### V. RESULTS AND DISCUSSIONS

For this study researcher collected sample data records of students. This information is stored in comma separated format which will be helpful for further processing. This dataset contains total 188 instances and 12 attributes which is shown in Fig.2. In this figure major parameters (aptitude, intelligence and interest) considered for career selection are shown below and are indicated by G1 to G9, NVTI and interest respectively.

Relation: training											
No.	1: G1 Nominal	2: G2 Nominal	3: G3 Nominal	4: G4 Nominal	5: G5 Nominal	6: G6 Nominal	7: G7 Nominal	8: G8 Nominal	9: G9 Nominal	10: NVTI Nominal	11: Interest Nominal
1	good	vgood	exell	exell	good	exell	exell	exell	exell	above_...	Commerce
2	medium	medium	exell	exell	good	exell	exell	exell	exell	above_...	Commerce
3	medium	weak	medium	good	medium	exell	good	exell	medium	above_...	Engineering
4	weak	medium	vgood	good	good	exell	weak	exell	good	above_...	Commerce
5	weak	weak	medium	good	good	exell	vgood	exell	weak	average	Medical
6	weak	weak	medium	medium	medium	weak	weak	weak	weak	average	Fine Arts
7	medium	exell	exell	best	good	exell	good	exell	medium	above_...	Arts
8	weak	vgood	good	exell	medium	good	medium	exell	weak	below_...	Commerce
9	medium	medium	medium	weak	medium	good	medium	exell	weak	below_...	Engineering
10	good	vgood	exell	good	good	exell	good	exell	good	above_...	Medical

Fig.2 Dataset in comma separated format

The rules generated by different classifier (FURIA, JRIP, PART and PRISM) are shown in Fig. 3, 5, 7 and 8 respectively.

```
(Interest = Engineering) => Option=science (CF = 0.53)
(G9 = exell) => Option=science (CF = 0.63)
(Interest = Commerce) and (G6 = good) => Option=commerce (CF = 0.8)
(Interest = Commerce) and (G3 = good) => Option=commerce (CF = 0.77)
```

Fig.3 Rules using FURIA with 2 fold

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.466	0.070	0.810	0.466	0.591	0.464	0.664	0.611	science
	0.938	0.047	0.652	0.938	0.769	0.759	0.986	0.782	commerce
	0.870	0.120	0.702	0.870	0.777	0.701	0.903	0.772	arts
	0.950	0.119	0.487	0.950	0.644	0.632	0.929	0.537	medical
	0.739	0.000	1.000	0.739	0.850	0.845	0.928	0.875	engg
	0.714	0.000	1.000	0.714	0.833	0.841	0.850	0.725	ITI
	1.000	0.011	0.600	1.000	0.750	0.770	0.995	0.600	farts
Weighted Avg.	0.707	0.073	0.763	0.707	0.701	0.630	0.823	0.693	

Fig.4 FURIA – Detailed Accuracy by Class

```
(Interest = Fine Arts) and (G8 = medium) => Option=farts (5.0/2.0)
(NVTI = below_avg) and (Interest = Engineering) and (G3 = good) => Option=ITI (3.0/0.0)
(G3 = medium) and (Interest = Commerce) and (NVTI = below_avg) => Option=ITI (2.0/0.0)
(Interest = Commerce) and (G6 = good) => Option=commerce (7.0/0.0)
(Interest = Commerce) and (G3 = good) => Option=commerce (3.0/0.0)
(Interest = Commerce) and (G1 = weak) and (G6 = exell) => Option=commerce (5.0/1.0)
(Interest = Medical) and (G5 = good) => Option=medical (28.0/11.0)
(Interest = Engineering) and (G5 = good) => Option=engg (29.0/14.0)
(Interest = Arts) and (G9 = weak) => Option=arts (32.0/6.0)
(Interest = Fine Arts) and (G8 = weak) => Option=arts (7.0/0.0)
=> Option=science (67.0/21.0)
```

Fig.5 Rules using JRIP with 10 fold.

=== Confusion Matrix ===

a	b	c	d	e	f	g	<-- classified as
34	7	14	16	0	0	2	a = science
0	15	1	0	0	0	0	b = commerce
1	1	40	4	0	0	0	c = arts
0	0	1	19	0	0	0	d = medical
5	0	1	0	17	0	0	e = engg
2	0	0	0	0	5	0	f = ITI
0	0	0	0	0	0	3	g = farts

Fig.6 Confusion Matrix using FURIA Classifier

```
Interest = Engineering AND
NVTI = below_avg AND
G3 = medium: science (4.0/1.0)

Interest = Engineering AND
NVTI = above_avg AND
G6 = exell: engg (14.0/6.0)

Interest = Engineering AND
NVTI = average AND
G1 = medium: science (15.0/6.0)
```

Fig.7 Rules using PART

```
If G3 = good
and G4 = good
and G5 = medium then commerce

If Interest = Arts
and G1 = exell then arts

If Interest = Arts
and G2 = medium
and G7 = medium then arts
```

Fig. 8 Rules using PRISM

Summary output obtained on collected dataset, using Weka’s classification algorithms is displayed in the following table I and table II.

TABLE I Classification Details using weka

Name of Classifier	Correctly classified instances (%)	Kappa statistic	Mean absolute error (MAE)	Root mean squared error (RMSE)	Relative absolute error (%)	Root relative squared error (%)	Coverage of cases
FURIA	70.74	0.632	0.0959	0.284	44.26	86.56	72.87
JRIP	69.68	0.59	0.131	0.256	60.55	78.00	97.34
PART	71.80	0.622	0.113	0.238	52.43	72.58	99.46
PRISM	100.00	1.00	0.0	0.0	0.0	0.0	100.00
ID3	100.00	1.00	0.0	0.0	0.0	0.0	100.00
J48	74.46	0.66	0.103	0.227	47.61	69.16	98.93
IB1	100.00	1.00	0.0	0.0	0.0	0.0	100.00

Above table describes correctly classified instances, kappa statistics and different error details.

Accuracy calculated using correctly classified instances divided by total instances. Kappa statistics is a measure of agreement between two individuals.

Mean absolute error (MAE) is a quantity used to measure how close forecasts or predictions are to the eventual outcomes. It measures average magnitude of the error in a set of forecasts, without considering their direction.

Root mean squared error (RMSE) is a good measure of the model's accuracy. RMSE is a quadratic scoring rule which measures the average magnitude of the error.

The MAE and the RMSE can be used together to diagnose the variation in the errors in a set of forecasts. The RMSE will always be larger or equal to the MAE; the greater difference between them, the greater the variance in the individual errors in the sample. If the RMSE=MAE, then all the errors are of same magnitude.[5]. Root relative squared error (RRSE) is the average of the actual values. Relative absolute error (RAE) is similar to the relative squared error.

TABLE II Accuracy in terms of weighted Average

Name of Classifier	TP Rate	FP Rate	Precision	Recall	F-measure	MCC	ROC Area	PRC Area
FURIA	0.707	0.073	0.763	0.707	0.701	0.630	0.823	0.693
JRIP	0.686	0.096	0.703	0.686	0.685	0.580	0.858	0.630
PART	0.718	0.113	0.708	0.718	0.709	0.605	0.900	0.725
PRISM	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000
ID3	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000
J48	0.745	0.093	0.755	0.745	0.742	0.654	0.919	0.759
IB1	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000

From above tables it is seen that classifier Prism, ID3 and IB1 gives 100% accuracy. These three classification algorithms gives best result as compared with other classifiers. TP rate means true positives rate and is proportion of positive cases that were correctly classified. It is calculated as correctly classified instances divided by total instances. From Fig.6 TP rate for class a =  $34 / (34+7+14+16+2+0+0) = 0.465$ .

FP is false positive rate and is the proportion of negative cases that was incorrectly classified as positive.

Precision is the proportion of the predicted positive cases that was correct.

Precision is calculated as correctly classified instances divided by total predicted instances.

Precision for class a =  $34 / (34+1+5+2) = 0.809$  and for class b is =  $15 / (15+7+1) = 0.65$ . Recall and TP rate are same.

F-measure is a measure of tests accuracy and is calculated using following formulae.

$$\text{F-measure} = (2 * \text{TP rate} * \text{Precision}) / (\text{TP rate} + \text{Precision})$$

$$\text{For class 'a' F-measure} = (2 * 0.465 * 0.809) / (0.465 + 0.809) = 0.591.$$

ROC (Receiver operating Characteristic area): ROC curve is given by TP and FP rate. Area under ROC curve is a method of measuring performance of ROC curve. If area under ROC is 1 then the prediction is perfect, and if it is 0.5 then prediction is random. From Fig. 4 it is seen that the best prediction is for class g followed by b, d and e and weak prediction is for class a.

Weighted average is an average in which each quantity is assigned with a weight used to calculate average. These weighing the relative importance of each quantity on the average.

## VI. CONCLUSIONS

Through this research paper, researcher explained framework design as well as working of the expert system useful for career selection. Data mining technique using different classification algorithms are useful for selection of career stream. Those are applied on collected dataset using Weka tool. Classification algorithms FURIA, JRIP, PART, PRISM, ID3, J48 and IB1 are applied and performance is observed. Machine learning algorithms viz; ID3, PRISM and PART gives 100% accuracy in classification along with rules. The details regarding classification accuracy, different errors are described in different tables. Thus classification is a good option for career selection process.

## ACKNOWLEDGEMENT

Researcher wish to acknowledge Mr. A.P. Shetbale (Counsellor, Bhilawadi Shikshan Sanstha, Bhilawadi), Mr. D.R. Thanekar (Counsellor, Institute of Vocational Guidance, Kolhapur) for providing information about career related data records of students and career selection process.

## References

1. G. Arulmani, "Career Psychology and Career Counselling: Core concepts for a Work and Education curriculum." pp. 1-21, 2005.
2. W1, "Description of WEKA (Java-implemented machine learning tool) Purpose:".
3. Career, "Career - Wikipedia, the free encyclopedia.".
4. Aptitude, "Aptitude and Ability Tests," 2015. [Online]. Available: <http://www.psychometric-success.com/psychometric-tests/psychometric-aptitude-tests.htm>. [Accessed: 29-Oct-2015].
5. W2, "Weka Data Analysis F-measure From the Weka Primer:" pp. 1-3.

## AUTHOR(S) PROFILE



**M.L. Waghmode**, received the M.C.A. degree in engineering faculty and M.B.A (HR) degree in Management. Presently working as Assistant Professor in BVDU, IMRDA Sangli.



**Dr. P.P. Jamsandekar**, received MCA degree in science faculty and Ph.D. in Computer Science. Presently working as Professor and Head of Department, Computer Applications, BVDU, IMRDA, Sangli, Maharashtra, India.