Careerchase – Career Guidance System using CBR and Rule Base
Jayashree Hajgude 1, Ayushi Gaur 2, A. P. Shreyas 3, Drishti Gurdasani 4, Mohit Changelani 5
Assistant Professor 1, BE Student 2, 3, 4, 5
Department of Information Technology
V.E.S. Institute of Technology, Mumbai, India

Abstract:
Career-related confusions are a serious issue among students these days. Choosing a right course in formative years is very important decision as one’s future depends on this one decision. Any student by himself is not mature enough to take right decision in his early life. Selecting wrong courses means mismatch between student aptitude, capability and personal interest. Faculty or parents have neither the required knowledge nor experience. Since there is no other reliable source generally available that can guide the student towards the most suitable direction, the career guidance system guides and helps students to choose a career by educating them on the basis of six key elements: academic background, hobbies, skills, income, IQ and personal interest. This proposed system helps students in Maharashtra studied under 10+2 education system to choose a career after undergoing a number of tests and also serves as a complementary tool for career guide and counselors. The following scripting languages are employed: PHP, MySQL, HTML, JavaScript and CSS.

Keywords: Career guidance, Case Based Reasoning, CBR, Data Mining, Rule based engine

I. INTRODUCTION
Guidance is a term sometimes used broadly to refer to advising or helping an individual with any kind of educational, vocational or personal problem. [14] It can also be referred to as a service provided by the particular school to help young persons in making clever decision and changes so as to develop their potentials as an individual and a contributing member of the society. Guidance activities are usually associated with educational professionals known as counselors as well as the involvement of parents, relatives, teachers, administrators, other educational specialists, spiritual leaders etc. The meaning is so broad that it does not deal with education alone but also aspects of life that affects an individual, therefore we can say that guidance is also a process of helping a person to realize and grow his/her vocational, educational, and psychological potentials and also achieving an the best level of individual happiness and societal usefulness. [2][10]. Provision of career counselling services are one of the main factors furnishing students’ academic success. [15] Main feature of student utilities is to give them the course best suited for their future and matches his attitude and attribute. Students choose particular courses of study because of perceived job opportunities, their interests and the likely future developments at the time of his completion of course. Problems arise if a student is not interested in the course or if career is not perfectly matching the student’s capability. Student counselling must include opinion on personal interests, handling interpersonal relation, qualities of learning strategies and also the attitude and aptitude. [2] Normally this activity is provided by counsellors or advisors who have lots of experience in the organisation. But with growing number of students and number of choices, and the amount of work on these advisors who are not able to handle the situation, the faculty of education institutions at higher secondary level does not have sufficient knowledge and experience of the courses and programmes except that of education. They also do not have time to counsel their students due to workload. Hence it is desirable to have some form of intelligent recommendation tools that needed to be developed to help them in the process of admissions. This problem determined the inspiration for this research and for developing the career guidance system to help students and staff. This system is recommended to be used to offer counselling for an entrant to a course to cope up between his capabilities and future requirements. The primary purpose of this research article is to discover the influence of career discovery. The module is a computer- assisted program for career decision making on the basis of marks and choice. The targeted population for this study was 10+2 passed students eager to join various fields like engineering, medical, commerce, arts etc. while for testing the system reliability students who have already made the choice were considered. A sample of this accessible population was drawn through random sampling procedures. [15]

II. NEED FOR THE SYSTEM
Students are novices with respect to career decisions, because they have no experiences which help them to evaluate their own vocational orientation. Experts tend to be over-specific and cannot take a "naive" point of view. Thus, an ideal system supporting vocational guidance should reduce and transform inputs from experts and generate results easily understood by students. [1]

This system solves the following problems:
1. Unavailable human counselors in some schools. This has resulting in many students making wrong career choices and then regretting for the rest of their lives.[14]
2. Where the counselors are available, there are mostly not enough to cater for all the students.
3. Good counselors are extremely expensive to hire and not every class of students can afford them.
4. Real life counselors are only available to students at working hours of the day, but an online career counselor is available at any time and just a click away.[14]
5. The education system and its offered courses vary from region to region. Hence, there exists no online system catering to the Maharashtra Education System.

The counselor’s role is to facilitate the client’s work in ways that respect the client’s values, personal resources and capacity for self-determination. It has become absolutely necessary because:

1. The internet has become a place where students (youths) visit often to interact (social network), shop (online shopping), be entertained (online video, music and radio services) etc. It is only proper that the internet is also used for something as productive and as important as interacting with an information system that helps them choose the best career possible for them.
2. If properly designed and implemented, an online career counselor can be more effective and accessible than a real life counselor.
3. It can be used to complement real life counselors; it can serve as a tool used by them.

III. REQUIREMENT FOR CAREER GUIDANCE SYSTEM

The system is restricted for a certain geographical area. Culture impacts the career decisions of students widely. The system is trained over data acquired from the students of a restricted area i.e. Maharashtra and so it might give output that is more relatable and acceptable to students belonging to similar area and culture. Also the education system followed by the students has to be 10+2. The machine is trained through the data sets which are collected in the form of questionnaires from the people of only one state of India. Limited number of datasets is being used. The attributes in the datasets are region-specific.

IV. OBJECTIVES

1. To know the process of career selection. [10]
2. To explore the problems encountered by the existing manual system
3. To identify factors affecting career selection. [10]
4. To work out the requirements of Career Guidance Systems for Educational Guidance.
5. To design a web based career guidance system that will improve upon the existing manual/human career guide. [14]

V. LITERATURE SURVEY

“Automated Career Counseling System for Students is using CBR and J48” by Maha Nawaz, Anum Adnan, Unsa Tariq, Jannat Fatima Salman, Rabia Asjad, and Maria Tamoor presents an automated system that mimics a one-to-one meeting with a professional career counselor. System is designed in a way that it takes inputs from the user, matches it with the training data and yields an output. The system’s fabrication involves more than one algorithm from Weka, a collection of Machine Learning algorithms for data mining tasks also the algorithm involved CBR of hamming distance with Manhattan distance to calculate the output as well as to refine the accuracy of the results. In this system, Decision Tree J48 is employed through Weka, to model the classification process for proposed majors. [7][10]. “Design of an online Expert System for Career Guidance” is an IEEE paper by S. Saraswathi, focuses on the construction of an online Expert System which guides the students for the selection of their undergraduate courses after the completion of higher secondary school education. The system is online in the sense that it will provide up-to-date information to the students. This information is acquired from web pages using pattern matching and jSoup parsing technique and the knowledge-base is constructed automatically without manual efforts. Rules are framed and an inference engine is developed which makes the Expert System. The constructed knowledge-base can be queried with domain related queries and the Expert System provides the most relevant details for the query. [8][10]. Dr. Abdullah Al-Ghamdi proposes a system that aims to assist postgraduate students of Computer Science (CS) major in King Abdulaziz University (KAU) to select the suitable courses during their postgraduate program. The proposed system enables the students to select and get a plan to each semester without needing to consult advisors. The architecture of the PAS model consists of four components: Knowledge Base (KB), Processing Unit, User Interface, and scheduler. The KB component consists of Database (DB) and Rule Base (RB). The system use the DB to store and retrieve the information about courses, students, prerequisites, core and elective subjects, sub-fields and the advising related information. The RB contains the rules and conditions that builds the PAS system. [13]. “A Proposed Decision Support System/Expert System for Guiding Fresh Students in Selecting a Faculty in Gomal University” by Muhammad Zaheer Aslam presents the design and development of a proposed rule based Decision Support System that helps students in selecting the best suitable faculty/major decision while taking admission in Gomal University. They designed a model using visual basic for testing and measuring the student’s capabilities like intelligence, understanding, comprehension, mathematical concepts his/her past academic record, intelligence level. This DSS identify the most suitable faculty or major for the student based on his abilities and capabilities extracted from the test module results. They used CLIPS language to store knowledge base. Rules can be made more customized and more criteria may be added to it for more data mined results. [16][10].

M. Ayman Al Ahmar, has developed a prototype rule-based Expert System with Object-Oriented (OO) modelling techniques for guiding high school students in selecting suitable undergraduate university majors. The system has a graphical user interface with simple menus. The architecture used in the design of the system resulted in a successful software system that is easy to maintain, modify, and extend. [1][8]. iAdvice is a Career Advisory expert system designed by Chathra Hendahewa et.al.to guide students for faculty of B.Sc. IT students of Moratuwa University, engaged in their higher education to determine their career paths and to select their course subjects to be in-line with their career goals. The System consists of three components viz; knowledge base, Inference Engine and user interface. This expert system uses features such as reasoning ability, providing explanations, alternative solutions, uncertainty and probability measures, questioning ability and also forward chaining, backward chaining and rule based inference in designing expert system. Past examination performance, student preferences and skills, industry alignment with subjects, are the main factors considered by a human expert in providing career guidance. [9][10].
VI. FRAMEWORK

CASE BASED REASONING

CBR is a computer technique, which combines the knowledge-based support philosophy with a simulation of human reasoning when past experience is used, i.e. mentally searching for similar situations happened in the past and reusing the experience gained in those situations. The concept of case based reasoning is founded on the idea of using explicit, documented experiences to solve new problems. The decision-maker uses previous explicit experiences, called cases, to help him solve a present problem. He retrieves the appropriate cases from a larger set of cases. The similarities between a present problem and the retrieved case are the basis for the latter’s selection.

\[ \text{sim}(x_1, x_2) = 0 \text{ corresponds to disim}(x_1, x_2) = 1 \]

The concept of similarity and dissimilarity comes into picture while calculating which cases match the given input. The distance between cases is calculated and the value is expressed as similarity or dissimilarity of the two cases. Note, that similarity and dissimilarity are opposites. Considering binary values \( \text{sim}(x_1, x_2) = 0 \) corresponds to \( \text{disim}(x_1, x_2) = 1 \)

The distance can be calculated using any of the following techniques:
1. Manhattan distance
2. Hamming distance
3. Euclidean distance

Advantages of CBR:
Case-based reasoning systems are an alternative in many situations to rule-based systems. In many domains and processes, referring to cases as a means of reasoning can be an advantage due to the nature of this type of problem solving. One of the most time consuming aspects when developing a rule-based system is the knowledge acquisition task. Acquiring domain specific information and converting it into some formal representation can be a huge task and, in some situations, especially less understood domains the formalization of the knowledge cannot be done at all. Case-based systems usually require significantly less knowledge acquisition as it involves collecting a set of past experiences without the added necessity of extracting a formal domain model from these cases. In many domains, there are insufficient cases to extract a domain model. There is another benefit to CBR, which is that a system can be created with a small, or limited, amount of experience and incrementally developed, adding more cases to the case base as they become available.

RULE BASED SYSTEM

The rule based system consists of three main components:
1. rule-base (permanent data)
2. an inference engine (process)
3. and a workspace or working memory (temporary data)

Knowledge is stored as rules in the rule-base. (Also known as the knowledge base)

Rules are of the form:
If some condition THEN some action

For instance, consider the following rules written in English Language:

Rule#1:
IF: the student has taken Elective subject E1
THEN: the student is inclined towards the following career options \{C1, C2, C5, ...\}

Rule#2:
IF: the family income of the student is in the range R1[.....lakhs]
THEN: the student is inclined towards the following career option \{C2, C6, ...\}

Rule#3:
IF: the student has the following hobbies & skills
THEN: he student is inclined towards the following career option \{C4, C5, C6, C8, ...\}

A rule is represented as a query stored in database, along with associated attributes like rule name, rule description and rule priority and 'where' clause of query contains condition and then part of the rule.

Advantages of Rule Based System: [8]
1. The principle advantage of production rules is notational convenience - it’s easy to express suitable pieces of knowledge in this way.
2. This would seem to be a purely declarative form of knowledge representation. One gathers pieces of knowledge about a particular subject, and puts them into a rule base. One doesn’t bother about when or how or in which sequence the rules are used; the production system can deal with that.
3. When one wishes to expand the knowledge, one just adds more rules at the end of the rule base.
4. The rules themselves are very easy to understand, and for someone (who is expert in the specific subject) the system is concerned with) to criticize and improve.
VII. MODULES OF THE DEVELOPED SYSTEM

SYSTEM FLOWCHART
Start [11]
Step 1: If (student is registered)
Then Login Input student id, password else signup Input student details endif
Step 2: Select stream
Step 3.1: Fill marks of compulsory subjects
Step 3.2: Select optional subject and fill its Marks
Step 4: Take IQ test
Step 5: Take interest quiz [2]
Step 6: Choose hobbies and skills
Step 7: Fill in family income
Step 8: Give result and explanation
Step 8.1: If (user chooses “in short”)
   Then show only final compiled output
   Else show detailed stepwise output endif
Step 9: Ask user to select a profession from the Suggested output and store it in Database
Stop

SYSTEM ARCHITECTURE

Figure 2. System architecture of careerchase

STREAMS AND COURSES

<table>
<thead>
<tr>
<th>Science</th>
<th>Commerce</th>
<th>Arts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Compulsory</td>
<td>Compulsory</td>
<td>Compulsory</td>
</tr>
<tr>
<td>Physics</td>
<td>Accountancy</td>
<td>History</td>
</tr>
<tr>
<td>Chemistry</td>
<td>Economics</td>
<td>Geography</td>
</tr>
<tr>
<td>Maths</td>
<td>Business Studies</td>
<td>Political Science</td>
</tr>
<tr>
<td>Optional</td>
<td>Maths</td>
<td>English</td>
</tr>
<tr>
<td>Biology</td>
<td>English</td>
<td>Optional</td>
</tr>
<tr>
<td>IT</td>
<td>Optional</td>
<td>Economics</td>
</tr>
<tr>
<td>Electronics</td>
<td>Entrepreneurship</td>
<td>Psychology</td>
</tr>
<tr>
<td>Computer Science</td>
<td>Physical Education</td>
<td>Sociology</td>
</tr>
<tr>
<td>Languages</td>
<td>Fine arts</td>
<td>Dance</td>
</tr>
<tr>
<td>Engineering (PCM)</td>
<td>Music</td>
<td>Music</td>
</tr>
<tr>
<td>PCB</td>
<td>Home Science</td>
<td>Fine Arts</td>
</tr>
<tr>
<td>MBBS</td>
<td>Commercial Art</td>
<td>Multimedia</td>
</tr>
<tr>
<td>BDS</td>
<td>Fashion</td>
<td>Fashion Designing</td>
</tr>
<tr>
<td>Veterinary Science</td>
<td>CA</td>
<td>Legal Studies</td>
</tr>
<tr>
<td>B.Pharmacy</td>
<td>BCom</td>
<td>Graphic Designing</td>
</tr>
<tr>
<td>Agriculture</td>
<td></td>
<td>Physical Education</td>
</tr>
<tr>
<td>Biotechnology</td>
<td></td>
<td>Entrepreneurship</td>
</tr>
<tr>
<td>BSc</td>
<td></td>
<td>Philosophy</td>
</tr>
</tbody>
</table>

Common Courses

| LLB | BFIA | Graphic Designing | BFA |
| BBA | BSW | Fire Sciences | Journalism |
| Architecture | Advertising | Psychology | B.English |
| NDA | BMS | Interior Designing | Event Management |
| HM | B.Ed | BA | Philosophy |
| Travel & Tourism | Aviation | Fashion Designing |

Figure 3. Streams after 10th, their corresponding compulsory and optional subjects, followed by their specific graduate courses and common courses
The user communicates with the CareerChase system through the user interface. It allows the user to query the system and supply information. The aim is to provide the same form of communication provided by a counsellor. With the help of user interface, necessary queries will be collected from the user. These queries are mapped with the rules from expert system. The queries will be on the region from which the student comes, which stream the student opted for, which branch the student prefers, fees he can pay, 10th and 12th percentage. All these queries will be the input for our domain. [8]

Figure. 4. Input of name, gender (personal details), SSC and HSC percentage

VIII. RESULT COMPUTATION

1. Based on academic marks (CBR)[12]
   Uses Case Based Reasoning to compare the user’s case with the existing cases, by comparing the marks obtained in subjects of the user with the stored database of students marks, to find the most nearest case. This system uses Manhattan distance to compute CBR.

2. Based on overall percentage (CBR)
   The student records having the same percentage ranges in 10th and 12th of that of the user are retrieved from the student’s database.

3. Based on elective subject (CBR)
   The Fields chosen by the students in the database, who had the same elective subject, are given as output.

Figure.5. Selection of elective subject in science stream

4. Based on interests (RB) [17]
   Interest quiz comprises of a set of 96 questions having answers in the form of {Not interested, little bit interested, somewhat interested, very interested}. [6] There are clusters of 6 questions pointing towards a set of fields. The quantitative interest of the user is known from this test. [19]

Figure.6. Interest Quiz and It’s Available Answer Options

5. Based on IQ (RB) [17][16]
   The IQ test has a database collection of 100 word problems and 100 image problems. The first stage of IQ test retrieves 10 word problems randomly followed by 10 image problems in the second stage from the database. The result of this IQ test is mapped with certain fields, which is given as output after undertaking the test. [4][19]

Figure.7. Word problems in iq test

Figure.8. Image Problems in Iq Test

6. Based on Income (RB)
   Ranges of Income are mapped with certain fields. The user chooses his family income range, and the output is generated accordingly. [19]

Figure.9. List of income ranges
7. Based on hobbies and personalities (RB) [17]
A large set of hobbies and personalities are listed down. The user can choose more than one hobby/personality from the list, and the mapped fields will be given as output. [19]

Figure.10. Hobbies and skill set to choose from

The fields suggested by each factor are collected upon one another and the the fields with maximum factor support are shown as results, refer figure XII. [10][11][18]

Figure.11. Storage of Result per Factor

OUTPUT TO USER FROM USER INTERFACE
In this module the user will be asked to view concise or descriptive result. If the user chooses concise result, he will only be shown top 5 career choices, refer to figure XII and XIII. And if the user chooses descriptive result he will be shown a detailed derivation of the top 5 career choices based on all the factors, refer to figure XIV. [8]

Figure.12. Concise result (bar graph)

Figure.13. Concise result (tabular) with priority (as count)

RESULT STORAGE
The user will be asked to choose one field from the top 5 fields, and this result will be stored in Case Base for future computations. This way the system learns and updates.

Figure.14. Detailed Result of All Factors

Figure.15. Final Career Selection Step (From Concise Result)

VIII. FINDINGS AND DISCUSSION
During the course of the research, the manual career guidance system was critically studied and some shortcomings were seen, places of improvement were discovered and a space that an online system can fit into was exposed.[14] We saw that an online career guidance system: CareerChase will prove really helpful in our present day, nearly everything is done online these days, and the age group of those that this research scope covers spends a good number of their time on the web doing less productive things, it is imperative and necessary that they be encouraged to engage in productive activities such as determining the best career possible for them.

IX. LIMITATIONS
The system is restricted for a certain geographical area. Culture impacts the career decisions of students widely. The system is trained over data acquired from the students of a restricted area and so it might give output that is more relatable and acceptable to students belonging to similar area and culture. The machine is
trained through the datasets which are collected in the form of questionnaires from the people of only one state of a country. Limited number of datasets is being used. The attributes in the datasets are specific.

X. CONCLUSION

CareerChase is an innovative idea. The opportunities provided by this e-medium are immense and many students can make use of this medium to choose a career more appropriate to their skills. In today’s competitive and technology driven world, with innumerable options available, the student is generally confused on choosing the right or more suited career. To conclude, the objective of designing this system is to lend a helping hand to the students aiming for such a career. And also serve as a complementary tool for career guide and counselors. By using our system, one will be guided towards a career to pursue. CareerChase currently deals with guiding students in a direction that is right for them i.e. to select a proper career path depending upon the present skill sets and mental abilities. Our system does not address the student’s inability to make reasonable judgments as a whole but tries to motivate careful consideration of as many career options as possible. [1] Furthermore this study also measures student’s skill strengths, abilities and personality facets and recommends them with possible career choices by using RB and CBR. RB makes the system easy to understand, implement and expand. The system can be created with a small, or limited, amount of experience and incrementally developed, adding more cases to the case base as they become available with the help of CBR. [20] This system will take the highest probability to be recommended for student but it also depends on the student whether they want to choose other careers with next four highest probabilities.

XI. FUTURE WORK

It is suggested that the scope of this research be widened if the time and resources permit. Career guidance is a continuous and never ending exercise. It can prove very useful not only for pre university students but also for first degree graduates, masters’ students and even prospective doctorate students as well. It can also be expanded for other states in India, and even beyond India.

XII. REFERENCES


[18]. Mr. Aristos Constantinou & Dr. Maria P. Michailidis, “A study of factors that influence students’ choices for undergraduate studies”, European Career Guidance Counselling Conference (2011)

