International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2018 IJSRCSEIT | Volume 3 | Issue 3 | ISSN : 2456-3307

Employability Analysis using Data Mining

Shubham Kambli[1], Shravani Joshi[1], Gaurav Chauhan[1], Deepali Nayak[2]

[1]B.E. Student Department of Information Technology, Vidyalankar Institute of Technology, Mumbai, India

[2]Assistant Professor Department of Information Technology, Vidyalankar Institute of Technology, Mumbai, India

ABSTRACT

In today's competitive world, it's getting difficult to get a job of suitable profile. It may be because of automation, recession or maybe the students themselves are not knowledgeable enough for the work expected from a particular profile. It can also be that students aren't aware of what the requirements are to work in a particular domain. Our system has been created to help students in all such scenarios. From a few assessment tests, it can provide you with an employability factor which shows how much employable you are for a particular domain, and also recommend courses, certifications to work on your weaknesses and gain an edge over others in the field. It can also suggest jobs which suit your current overall profile. This system can be used by students to assess themselves, by placement cells to measure employability of their student, and also by companies to check employability potential of prospective employees.

Keywords : Data Mining , Employability Analysis, Recommendation

I. INTRODUCTION

The system for employability analysis has been around since many years. Companies, placement cells use academic records of the students to check if he/she is a good fit for a profile [1]. In recent years, many assessment tests such as AMCAT, e-Litmus have become popular. However we are trying to include Data mining capabilities to learn how employable graduate students from engineering background are. We also plan to integrate a recommender system to recommend training that students must undergo to become a more employable resource. The parameters for measuring employability will also be revised and accordingly students would be suggested with suitable training to target their desired job profiles. Companies would also be able to hire students that suit their particular job profile by viewing their profile as a whole [2]. So, overall it would be a win-win situation for both the students as well as the companies who would be hiring them.

II. EXISTING SYSTEM

Existing System for employability analysis only involves conducting tests that determine the credibility of a student only by means of aptitude tests and technical assessment tests. Also, when it comes to assessing this as well, existing systems only focus on of these factors individually and not as a whole. If one parameter exists, the other doesn't. Such systems lack a comprehensive analysis of data from multiple fields.

III. AIM & OBJECTIVE

The main aim of our project is to analyze how much employable a particular student from an engineering background is. In order to determine the employability of a particular student, we will be assessing their overall profile. Data Mining algorithms would be applied on the data obtained from the students and accordingly and they would be

suggested training based on whatever job profile they are trying to target.

The main objective of our project would be to provide employability as a quantifiable factor to all the students who are in search of a job. Also, we will be helping the students with training need analysis, i.e. students could assess themselves and determine what areas they lack in and work upon it. This would help the students gain a competitive edge over their peers who would not be using this system. Students, as a whole, can then nurture their existing talent as well as work upon their weaknesses, leading to overall personality development [3]. This, as a result would help us achieve the vision of Skill India mission which works on employing the youth in India.

IV. LITERATURE SURVEYED

Title of Paper, Journal name	Observations/ Process Flow	Challenges/ Limitations	Inferences / recommendations
Application of Data Mining in predicting placement of Students, IEEE	Academic Results, Programming Skills and Soft Skills of students were taken into consideration in order to determine what chances the students have in order to get placed and what companies they'll get placed in.	Data acquiring and cleansing being the major challenges here, this study is fairly accurate and does not face any major issues.	Academic details, Programming Skills, Soft Skills should be taken into consideration. Data cleaning must also be done [1].
A Placement Prediction System Using K-Nearest Neighbors Classifier	Strengths and weaknesses of students are identified in various fields in order to determine their probability of getting placed. Students could assess and determine the best of their abilities in order to determine the most suitable job profiles for them.	The study has been done only for IT field and its scope is therefore, limited to it. Further changes have to be done in order to make it suitable for other fields and streams as well since the scope is limited.	Strengths and weaknesses of students must be considered [2].
Mining Student's Data for Performance Prediction, IEEE	Academic and personal as well as social factors were used in order to determine student grades. Various attributes were considered in a discrete manner to predict the results.	Socio- economic conditions were found to have marginal effects in the performance. Hence, not much could be determined with the help of those.	Socio- economic factors have a marginal effect [4].
Academic Performance Predictors, IEEE	Academic, Socio- economic and personal parameters were used to predict student performance. Of these, it was found that the academic parameters are the ones which are highly influential in predicting the student	The results depend on what the objective and goal for prediction is and hence vary greatly as the conditions change. There might exist parameters which may not have been included but could	Personality characteristics should also be included [5].

	performance while Socio-economic and Personal parameters do count, although a bit insignificant as compared to the academic ones	act as good predictors.	
PPS - Placement Prediction System using Logistic Regression	There is a general trend by the companies to hire students based on particular skill sets. Different jobs for different fields are determined based on the particular skill sets as required by the companies.	Various skills that need to be developed for particular companies is determined by this model. Its scope is limited only to develop particular skills.	Different companies require different skill sets according to the job profile [6].
Prediction of Students Performance using Educational Data Mining	Various Academic and personal factors of the students are gathered to determine the academically weak students and devising methods to prevent them from getting a drop.	It determines chances of failure of a student and not success and doesn't suggest any plan to improve performance.	Gap Analysis should be included [8].
Optimal selection of factors using Genetic Algorithms and Neural Networks for the prediction of students' academic performance	It basically takes into consideration a lot of relevant factors most of them being really significant broadly divided as intellectuals, ability to study and personality. These factors are used for determining various things such as probability of getting a drop, getting a placement, etc.	Takes into consideration a lot of factors, many of them being pretty insignificant. To Optimize the solution numbers of factors have been reduced to 8 from 39.	Generalize factors into categories and then use them accordingly [10].

V. PROPOSED SYSTEM

The legacy system for testing employability lacks newly found technologies. In our proposed solution, we are bringing the recent Data Mining advancements in this employment domain. With our solution — Data mining algorithms — we are quantifying the employability factor for the graduates. The parameters for measuring employability will also be revised. So, overall what we would be doing is providing students with a system that would help them assess themselves and help them identify their strengths and weaknesses for targeting a specific job profile. For developing the project, we will be using Hadoop, which is an open source programming framework that supports the processing and storage of extremely large data sets in a distributed computing environment [5]. As far as programming language is concerned, we will be using Python for it since it can be integrated with the Hadoop based framework that we have. When it comes to data mining algorithms, our primary focus will be on Naive Bayesian classifier which would be used for classifying the data that we have and finding suitable patterns which we can use to determine employability. Naive Bayesian classifier follows a

probabilistic approach and thus can be used for integrating, processing and classifying quantifiable data to get the results. Along with it, we also plan on using J48 and Random Forest classification algorithms. All such algorithms will classify the data and our focus will be on maximizing the accuracy of such data mining algorithms to classify the data in a proper manner.

VI. WORKING OF PROPOSED SYSTEM

In the previous section we have understood the various parts of the proposed system. It even clarifies about the significance of each of the parts. Now in this section, let us understand the working of the proposed system. This would enable us to understand how this proposed system can be effectively used to get rid of the problems in traditional method.

Project methodology involves collecting quantifiable data from students such as

1. Academic details
2. Internships done by them
3. Skills known
4. Personality assessment, etc

and assigning weights to all such factors in accordance with the contributions that they make to the overall performance of the student. Once all such factors have weighed in we calculate the overall employability for that person and accordingly suggest various approaches to develop the skills required for a desired position [6]. All such data would be integrated, preprocessed and classified with the help of data mining algorithms, while acquiring whatever knowledge and patterns we can from the existing data so that it could be used for analysis by the students. Students can themselves see what all factors are there that they need to work upon, what factor contributes in what way and how much for a particular job profile. Suppose a student is good for a particular profile and they don't know what profile they wish to have or what they wish to pursue then the system can accordingly suggest them with what they are good at or can be so that they could further strengthen themselves in accordance with that particular profile [7]. Students who are already aware and know what they wish to pursue can just check in on what factors weigh in and in what way, so that they could work upon those in order to acquire the desired position [8].

Our project would work in multiple iterations which would comprehensively be covering all the major stages that are involved in the project. Prototype model would help us better assess whatever mistakes that have been made in the previous iterations so that they could be rectified and improved upon in the further iterations. Feedbacks can also be received from the users so that they could specify their needs and the system could be customized to whatever the needs of the user are so that it could better serve its purpose [9]. Whatever functionalities the user deems missing could be added. Students would thus help improve the system with their valuable feedbacks as well as companies could suggest what their requirements are specifically which could be added in further iterations.

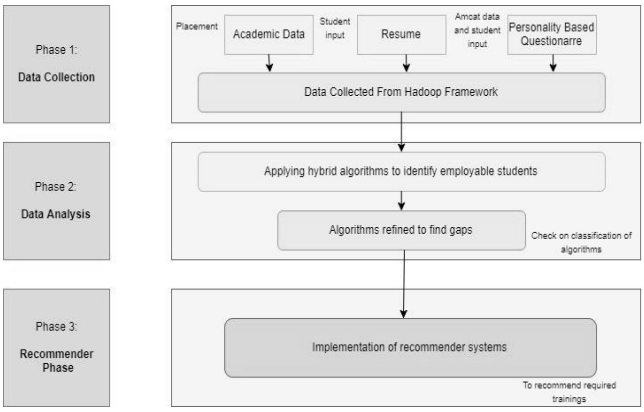

Figure 1. Architecture Diagram

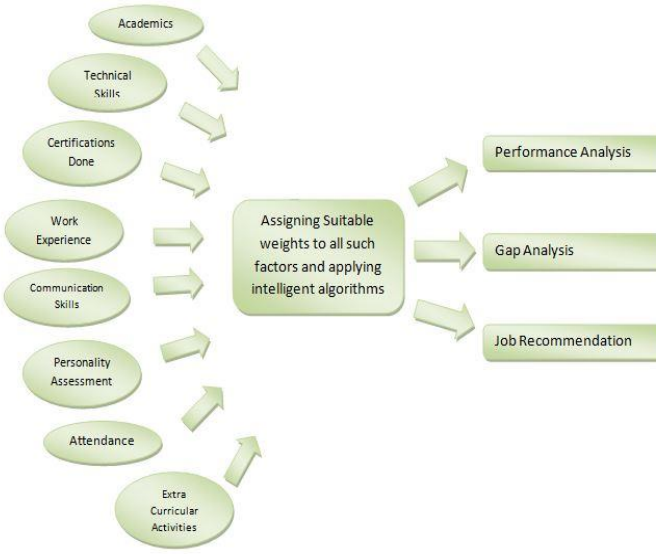

Figure 2. Parameters selected for analysis

The project will be broadly divided into 3 iterations:

1. Data Collection:

This stage would involve collecting data from the users and integrating it. Various data would be collected in the form of factors that we would need to assess students' employability. Factors that we would be considering for analyzing the employability of person include academics, internships/freelancing done, certifications done, personality assessment data, aptitude wise performance, technical skills acquired, attendance, etc [10].

2. Data Preprocessing and Analysis of data:

All these factors would be integrated and weighed in response to how they affect the employability of a person. The more the weight assigned to the factor, the more it would contribute in helping the person get employed. Before analysis can be done, all such data would be cleaned i.e. preprocessed so that things like missing values can be removed, redundant data can be removed, as well as all the inconsistencies would be resolved. Pattern analysis would also be done for determining the trends in hiring for a particular job profile so that weights can be assigned to the factors.

3. Recommending training to students after gap analysis:

Student data would then be judged on the basis of the database with the help of various data mining algorithms so that they could be recommended with what could be done for getting a job according to their desired job profile.

VII. CONCLUSION & FUTURE WORK

Engineering students who wish to assess themselves can use this system in order to check their employability.

Engineering graduates can use it to target specific companies that suit their needs. Engineering undergraduates can also use it to better utilize their time in order to get ready for the challenges that they may be facing in the future in order to get a job in this already competitive world. They would be able to get themselves ready in time for the future struggles that they may face. Companies may also use this system in order to pick up students that fit their needs.

Future scope for this project could be applying even more sophisticated algorithms which could be customized for this particular application. Machine learning capabilities could also be included for this method of finding employability of a person.

VIII. REFERENCES

[1]. Pratiyush Guleria and Manu Sood, "Predicting Student Placements using Bayesian Classification," presented at the Third International Conference on Image Infonnation Processing, Waknaghat, India, 2015.

[2]. Mansi Gera and Shivani Goel, "A Model for Predicting the Eligibility for Placement of Students Using Data Mining Technique," presented at the International Conference on

Computing, Communication and Automation (ICCCA2015), Noida, India, 2015.

[3]. Tripti Mishra and Dharmiinder Kumar, "Mining Students Data for Performance Prediction," presented at theFourth International Conference on Advanced Computing and Communication Technologies, Rohtak, India, 2014.

[4]. Jaime Ramirez Castillo and Aldabbagh Ghadah and Habib M. Fardoun," Towards improved student placement and preparation methods on IT post-secondary education," in Proceedings of the 2013 Federated Conference on Computer Science and Information Systems, Krako, Poland, 2013.

[5]. Cheng Lei and Kin Fun Li," Academic Performance Predictors," presented at the 29th International Conference on Advanced Information Networking and Applications Workshops, Gwangiu, South Korea, 2015.

[6]. Ajay Shiv Sharma and Swaraj Prince and Shubham Kapoor and Keshav Kumar," PPS - Placement Prediction System using Logistic Regression," presented at the IEEE International Conference on MOOC, Innovation and Technology in Education (MITE), Patiala, India, 2014.

[7]. Animesh Giri and M.Vignesh and V. Bhagavath and Bysani Pruthv and Naini Dubey," A Placement Prediction System Using K-Nearest Neighbors Classifier," presented at the Second International Conference on Cognitive Computing and Information Processing (CCIP), Mysore, India, 2016.

[8]. Karan Pruthi and Parteek Bhatia," Application of Data Mining in predicting placement of Students," presented at the International Conference on Green Computing and Internet of Things (ICGCloT), Noida, India, 2015.

[9]. Omar Augusto and Echegaray-Calderon and Dennis Barrios-Aranibar," Optimal selection of factors using Genetic Algorithms and Neural Networks for the prediction of students' academic performance," Computational Intelligence (LA-CCI),2015 Latin America Congress, Curitiba, Brazil, 2015.

[10]. Tismy Devasia and Vinushree T.P. and Vinayak Hegde," Prediction of Students Performance using Educational Data Mining," presented at the Data Mining and Advanced Computing (SAPIENCE), International Conference, Ernakulam, India, 2016.