Demand Analysis of CSE Graduates of Different Universities in Job Markets

Debjany Chakraborty, Md. Sabir Hossain, Mohammad Shamsul Arefin

Department of Computer Science and Engineering
Chittagong University of Engineering and Technology (CUET)
Chittagong-4349, Bangladesh
Email: debjany.cuetcse@gmail.com, sabir.cse@cuet.ac.bd, sarefin@cuet.ac.bd

***Abstract*— CSE is one of the most encouraging fields with high growth opportunities in the job sectors. Every year we are getting a large number of CSE graduates from public and private universities of Bangladesh as well as the whole world. Each of them tries to get a good job with better salary as early as possible. Several factors (university ranking, CGPA, programming skill, job experience etc.) are connected to get a good job. In this paper, we have analyzed CSE graduates of different universities in job sectors with the help of decision rule. We have collected and analyzed data from ten public and private universities of Bangladesh. We have generated graduate's individual rank and finally produced rank of ten universities. We have applied the Pearson Correlation formula to recommend salary and approximate time estimation to get a job. We evaluated our system several times with the necessary graduate's profile and got a satisfactory result.**

***Keywords*—CSE Graduates, Decision Rules, Pearson Correlation, Recommendation System.**

I. INTRODUCTION

Every year a large number of students complete their graduations from different private and public universities and try to get a good job as early as possible. The IT field in Bangladesh is growing at a high speed in recent years. Also, job opportunities are comparatively better in this industry. For this reason, Computer Science and Engineering (CSE) is becoming the most encouraging field with high growth opportunities. But there are several factors which are connected with getting a good job and a better salary. These are university name, CGPA, programming skill, job experience etc.

To analysis and observe the Computer Science & Engineering graduates in Bangladesh, we developed this framework. We have considered five topmost public and five topmost private universities for analysis purpose. We analyzed the graduate's profile like cumulative grade point average (CGPA), time difference between their graduation and joining of the first job, their joining salary and also present salary with their present job address, their programming skill and so on.

We have developed a system that can:

- Provide a ranking of the graduates from different universities
- Produce a rank of the universities we have considered
- Analyze reasons (CGPA, Programming Skills, etc.) behind getting expected jobs through decision rules
- Recommend job with an estimated salary and an estimation of time for finding a using Pearson Correlation Formula

The rest of the paper is organized as follows: Section II provides a brief review of related work. Section III discusses in details about the system architecture and design. Section IV presents implementation and performance evaluation. Section V concludes the paper and recommends the future directions of this work.

II. RELATED WORK

Qasem et al. [1] tried to use the data mining algorithm such as classification algorithms to measure the performance of the students in order to enhance the higher education system quality. They evaluated the student's data which might affect the student's performance in C++ courses. Their methodology was to implement a CRISP framework to carry out their research. They compared the results and found that the Decision tree C4.5 algorithm performed better than other algorithms.

Surjeet et al. [2] compared and analyzed the C4.5 algorithm, ID3 algorithm, and decision tree (CART) algorithms to predict the accomplishment of engineering students of the first year in examinations. The prediction was carried out in three separate classes for those who failed, promoted and passed.

Kabakchievaet al. [3] devised a system for predicting the performance of students and they applied some data mining algorithms such as Decision Tree, OneR Rule Learner, Neural Network and K-Nearest Neighbor. It was applied to the Bulgarian student's dataset. In this paper, they classified the problem into two class named weak and strong. This model achieved the best accuracy for prediction of strong class but good for predicting weak classes.

Dharaneeshwaran et al. [4] analyzed three separate processes for obtaining similarities to generate a recommendation. There are several mathematical techniques which might be used to obtain the similarity between items. After taking consideration on different parameters, they came

978-1-5386-9111-3/19/$31.00 ©2019 IEEE

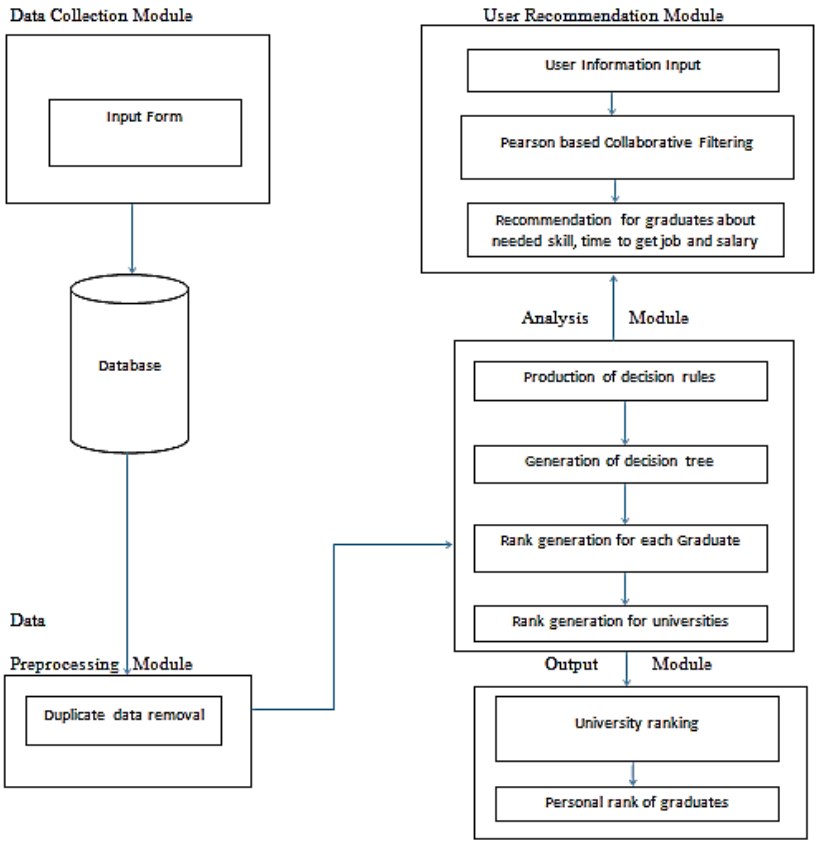

Fig 1. The system architecture of the proposed system

to the decision that Pearson's Correlation Coefficient gives better performance than all other collaborative algorithms, they have considered.

Sulaiman et al. [5] identified the optimal decision rule algorithms for predicting the student's performance in a computer programming course taken in 200 level based on their results in Physics and Mathematics in 100 level. They look into various techniques for computing item-item similarities and generating recommendations. They also revealed that previous knowledge of Mathematics and Physics both at ordinary level and 100 levels are essential determinants of student's performance in a computer programming course.

Sarwar et al. [6] analyzed various item-based recommendation algorithms. They observed various procedure to compute item-item similarities and various techniques for generating recommendations from them. At last, they evaluated results and compared with the basic k-nearest neighbor process. In this paper, they suggested that item-based algorithms give much better performance than user-based algorithms to the user.

A collaborative filtering algorithm based items recommendation technique using Pearson Correlation formula was proposed by Tomar et al. [7]. Pearson correlation calculation finds out the similarity measure of the products and recommends them according to the estimated similarity.

III. System Architecture and Design

The system architecture of our developed system is shown in Fig. 1. The system architecture of the Computer Science & Engineering Graduate Analysis system comprises five basic modules. They are 1) the data collection module, 2) the second module is a data preprocessing module, 3) the analysis module, 4) the output module, and 5) the recommendation. The function of the data preprocessing module is to remove duplicate data and also missing value. Database system stores all the information. The third module analyzes data with decision tree algorithm. The fourth module shows the ranking of university and graduates also. The fifth module recommends user salary and expected time to get a job.

A. Data Collection Module

We have collected our relevant data by launching a website. The data fields are name, university name, CGPA, known programming languages, graduation date, first job joining date, first job designation and address, present job designation and address, joining salary, present salary and comment about the job which is categorized into three consequent terms like: not satisfied, fully satisfied and apparently satisfied. All the data was stored in our database for further analysis.

B. Data Preprocessing Module

We collected data through a sign-up form using a website hosted. As some person submit the form more than once, some redundant data present in the database. We have removed the redundant data using SQL query in order to get the accurate result from the analysis module. Also, admin of the system has the power to update or delete any row in the database.

C. *Data Analyzing Module*

In this module [Fig. 2], at first, we generated a decision tree. For generating decision tree we need to calculate information gain. Here, we have calculated the information gain to select the attribute level. The attribute which holds the highest info gain will be the root node and lower values will follow. Here, the difference of joining date and graduation date gets the highest gain (0.652), which we have added as the root of our decision tree.

There is a time difference between graduation date and first job joining date, joining salary, increment, satisfactory level. We count the increment using (1).

Increment = (Present Salary-Joining Salary) / (Present date – Joining Date) (1)

We find out the time to get a Computer Science & Engineering graduates first job by checking out the difference of his / her graduation date and first job joining date. We also find out some rules for generating a decision tree for carrying out the analysis of these graduates information.

We have covered all possible paths existing in the collected dataset. We have generated 81 ranks of Computer Science & Engineering graduates using decision tree rules. Some samples decision tree is provided in Table I. We split the tree into four levels shown in Fig 3. We took the best attributes from the perspective, the difference of graduation date and first job joining date. We took three range of it like: less than six months, six to twelve months and the third one is greater than twelve months. The next level attribute is the joining salary which is also categorized by three different ranges such as salary less than twenty-five thousand, salary between twenty-five to forty thousand and last of all salary greater than forty thousand taka. The third level of the proposed system is the increment of the Computer Science & Engineering graduates. We included the increment of salary due to switching jobs. We have considered three possible categorizations of increments are less than five thousand taka, between five to ten thousand taka and greater than ten thousand taka.

In the second decision tree [Fig 4.], we have considered four attributes: the difference between graduation date and joining date, joining salary, CGPA which is categorized into three ranges provided in Table II. and the number of known programming languages.

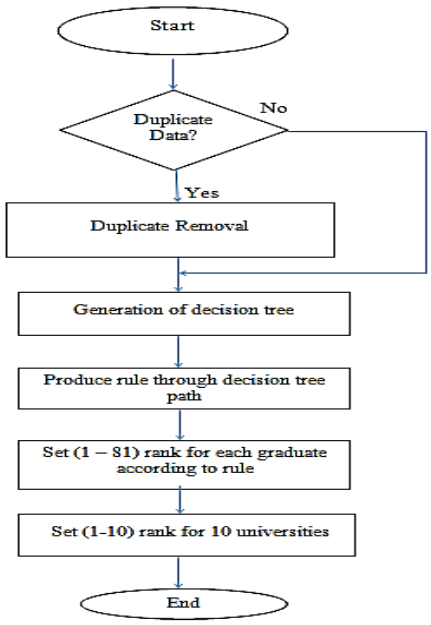

Fig 2. Flow graph for the recommendation module

D. *Output Module*

In Output Module, the user can see the university ranking of ten universities of Bangladesh in a bar graph [Fig. 7]. The graduates can see personal rank according to his/her university from the graph. We have set the rank between 1 to 81 for each graduate. Graduates will also get the recommendation of expected job expected salary and job acquisition time.

TABLE I. A SAMPLE OF DECISION TREES

Sl. No.	Rules Set 1	Rules Set 2	Rank
1.	<6 -> <25 -> <5 -> fully satisfied	<6 -> <25 -> 1 -> <=2	25
2.	<6 -> <25 -> <5 -> apparently satisfied	<6 -> <25 -> 1 -> <2-4	26
3.	<6 -> <25 -> <5 -> not satisfied	<6 -> <25 -> 1 -> >4	27
4.	<6 -> <25 -> 5-10 -> fully satisfied	<6 -> <25 -> 2 -> <=2	22
5.	<6 -> <25 -> 5-10 -> apparently satisfied	<6 -> <25 -> 2 -> <2-4	23

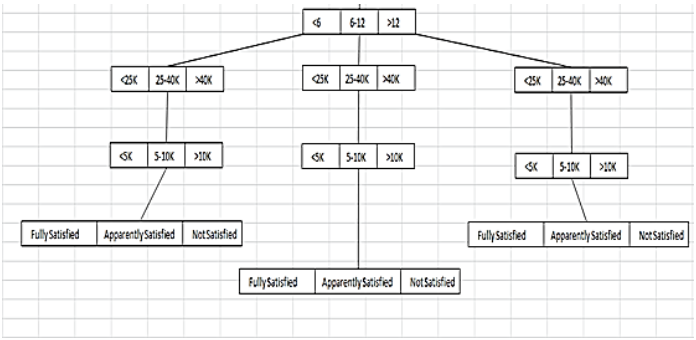

Fig 3. Symbolic representation of our proposed decision tree 1

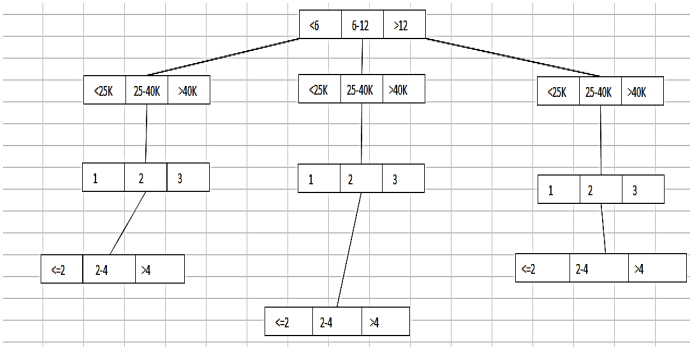

Fig 4. Symbolic representation of our proposed decision tree 2

E. Recommendation Module

In Recommendation Module [Fig. 5], we used Pearson Correlation Formula [8] for measuring the strength between variables and relationships.

Pearson's correlation coefficients for user A and U is:

$$w_{a,u} = \frac{\sum_{i=1}^{n}(r_{a,i}-\overrightarrow{r_a}).(r_{u,i}-\overrightarrow{r_u})}{\sqrt{\sum_{i=1}^{n}(r_{a,i}-\overrightarrow{r_a})}.\sqrt{\sum_{i=1}^{n}(r_{u,i}-\overrightarrow{r_u})^2}} \quad (2)$$

$$P_{a,i} = \overrightarrow{r_a} + \frac{\sum_{u=1}^{m}(r_{u,i}-\overrightarrow{r_u}).w_{a,u}}{\sum_{u=1}^{m}|w_{a,u}|} \quad (3)$$

Where a= the active user, u= another of the users of the system, n= the number of items that both the active user and all recommender users have rated, ra= is the average ratings of the active user ru is the average of another user u's ratings, w (a,u) = degree of correlation between user a and user u. We have calculated our collected data in different possible ranges. Such as CGPA in three ranges [Table II], joining salary in five ranges [Table III], the rank of the university in Table IV.

TABLE II. CGPA RANGE

Sl. No.	CGPA Range	Rating
1.	3.50-4.00	1
2.	3.21-3.49	2
3.	2.75-3.20	3

TABLE III. SALARY RANGE

Sl. No.	Salary Range	Rating
1.	65000-100000	1
2.	50000-64000	2
3.	30000-49000	3
4.	15000-29000	4
5.	5000-14000	5

TABLE IV. UNIVERSITY RANKING

Sl. No.	University name	Rating
1.	BUET	1
2.	CUET	5
3.	KUET	3
4.	RUET	2
5.	DU	4
6.	NSU	7
7.	BRAC	10
8.	AIUB	8
9.	AUST	4
10.	MIST	9

We have calculated the Pearson Correlation among the users to provide recommendation listed in Table V.

TABLE V. USER ITEM MATRIX

User	CGPA	Salary	University
User A	1	2	5
User B	4	3	6
User C	3	5	2
User D	4	-	3

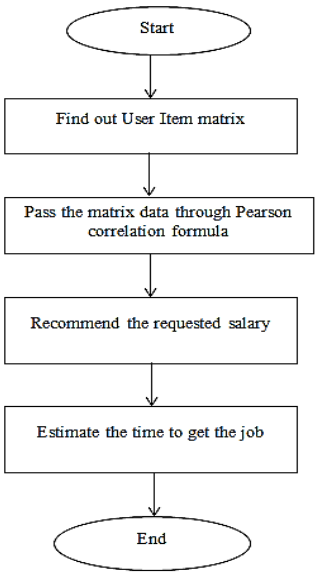

Fig 5. Flow graph for recommendation for a graduate

The average rating of the active user A, B and C are defined by rA, rC and rC respectively.

ra: (4+3)/2=3.5

rA: (1+5)/2 = 3

rB: (4+6)/2 = 5 r

C: (3+2)/2 = 2.5

The average of user C's ratings which is (3+2)/2 = 2.5

$$W_{a,d} = \frac{-2}{\sqrt{0.5}\sqrt{0.5}} = -1$$

$$w_{b,d} = \frac{-1}{\sqrt{2}\sqrt{.5}} = -1$$

$$w_{c,d} = \frac{0.5}{\sqrt{0.5}\sqrt{0.5}} = 1$$

The correlations between the other users and user D are calculated from the equations (2) and (3).

IV. IMPLEMENTATIONS AND EXPERIMENTS

In this section, the implementation details and performance evaluation of our proposed system are discussed.

A. Implementation

As we worked with a large number of professional and personal sensitive data of CSE graduates, the system has an authentication system for user access. We have collected data

from different universities by hosting a website. The salary recommended to a graduate and approximate time to get job is shown in Fig. 6. On clicking the graduates rank the following rank of each individual university BUET, KUET, RUET, CUET, DU, NSU, AIUB, MIST, AUST, BRAC will be shown in Fig. 7. We have analyzed the Graduate's profile and set rank for each graduate.

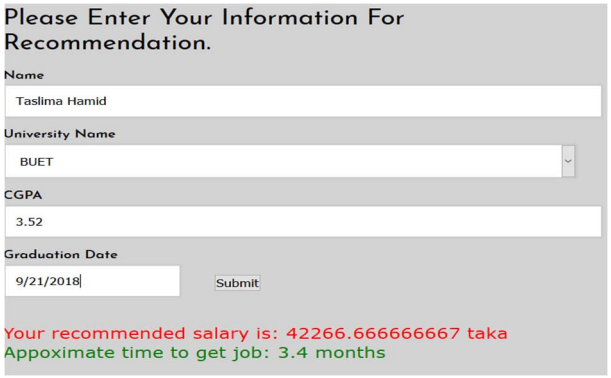

Fig 6. Recommendation for a graduate

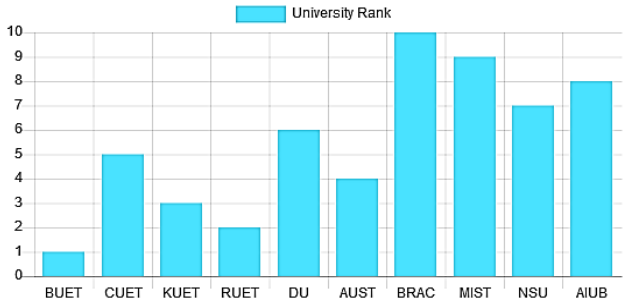

Fig 7. Visualization of university ranking

C. Performance Evaluation

We have tested the performance of our developed system using the Mean Absolute Percent Error metric. We analyzed collected data and produced a rank of the universities and generated an individual rank for the graduates. Based on rank and other factors, the system recommends estimated salary and approximate time for the graduates to get a job.

The MAPE (Mean Absolute Percent Error) calculates the error size of a system in percentage using (4).

$$\text{MAPE} = \frac{1}{n}\sum\frac{|Actual - Recommended|}{Actual} * 100 \qquad (4)$$

We measured the system efficiency [Table VI] and get the satisfactory result shown in Fig. 8. We have evaluated our proposed system for a different number of graduates starting from 10. Each time, we have increased the number by 10. We have tested salary recommendation and time estimation efficiency ten times. We observed that the efficiency increased with the increase in the graduate number shown in Fig. 9.

For example, in the case of BUET, the percentage of error of salary recommendation is 8.35% and efficiency of the proposed system is 91.64%. The percentage of error of estimation of time to get a job is 16.32% and proposed system efficiency is 83.67%.

TABLE VI. SYSTEM EFFICIENCY CALCULATION

Graduates Number	System Error	System Efficiency
10	34.90%	65.99%
30	10.98%	89.02%
40	8.08%	91.97%
50	6.02%	93.97%
60	5.63%	94.37%
70	4.10%	95.90%
80	3.11%	96.89%
90	3.02%	96.98%
100	3.01%	96.99%

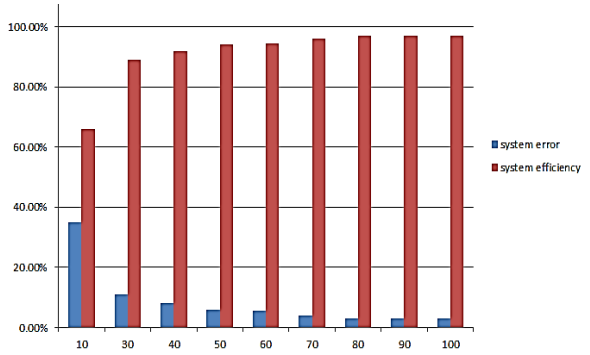

Fig 8. Graphical representation of system efficiency (analysis module)

V. CONCLUSION

In this paper, we have developed a system to analyze and observe the CSE graduates in Bangladesh. We have analyzed the graduate's profile and generated a rank for them as well as

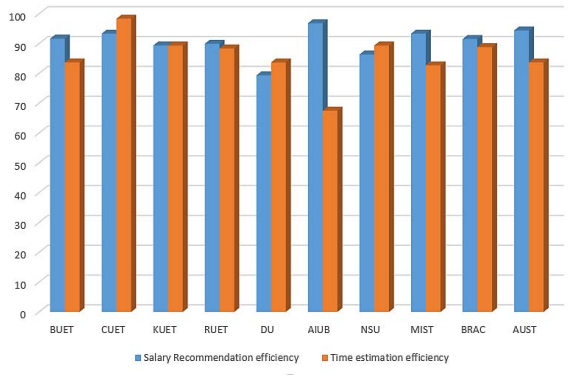

Fig 9: Graphical representation of system efficiency

their university. Then we have recommended graduate's salary at the starting of their career and estimated time limit to get the first job. The proposed system can visualize the real scenario of our IT sector. We have also pointed out the factors responsible for the failure of graduates, delay of improvement for their career. If university authority and graduates themselves at their student life concentrate on those issues, we can boldly propose that their career will be up to the mark on the aspect of Bangladesh.

We have analyzed only ten universities in Bangladesh. The more accurate result might be achieved by considering all universities with CSE and related departments.

REFERENCES

[1] Q. Radaideh, E. Shawakfa and M. Najjar, "Mining student data using decision trees," In *International Arab conference on information technology*, pp. 1-5, 2006.

[2] S. Yadav and S. Pal "Data mining: A prediction for performance improvement of engineering students using classification", *In World of computer science and information technology journal (WCSIT)*, vol. 2, no. 2, pp. 51-56, 2012.

[3] D. Kabakchieva, "Student performance prediction by using data mining classification algorithms," In *International journal of computer science and management research*, vol. 1, issue 4,pp 686-690, 2012.

[4] A. Dharaneeshwaran, "Calculating the user-item similarity using Pearson's and cosine correlation," In *International conference on trends in electronics and informatics (ICEI)*, pp. 42-49, 2017.

[5] A. Sulaiman, B. Akinbowale and B. Ronke, "Comparative analysis of decision tree algorithms for predicting undergraduate students' performance in computer programming," In *Journal of advances in scientific research & its application (JASRA)*, vol. 2, pp. 79 – 92, 2015.

[6] B. Sarwar, G. Karypis, J. Konstan,, J. Riedl, "Item-based collaborative filtering recommendation algorithm," In *Eleventh international conference on data engineering*, 1995.

[7] A.S. Tomar, "An implementation of Pearson correlation method for predicting items to a user in e-commerce," In *International journal for research science & technology*, pp 873-882, 2016.

[8] "Correlation coefficient: simple definition, formula, easy steps", https://www.statisticshowto.datasciencecentral.com/probability-and-statistics/correlation-coefficient-formula/ [Last access on Oct. 21, 2018].