Improving Students' Motivation to Study using Salary Prediction System

Pornthep Khongchai

Department of Computer Science

Faculty of Science and Technology

Thammasat University Pathumthani, Thailand
pornthep.khon@thammasat.net

Pokpong Songmuang

Department of Computer Science

Faculty of Science and Technology

Thammasat University Pathumthani, Thailand
pokpong@cs.tu.ac.th

Abstract— **This paper proposes a salary prediction system for increasing students' motivation in studying. A decision tree technique is used to generate a prediction model with seven features. We evaluated the system efficiency using 13,541 records of graduated student data in 10-fold cross validation. An accuracy result in overall is 41.39%. Furthermore, we exploit questionnaires to evaluate effectiveness of the system with 50 student samples. The result shows that the system can effectively boost students' motivation in studying and also show them a positive viewpoint of their future. At last, the sampling students answer the positive satisfaction in using the system since they found the system is easy to use and the prediction results are simple and comprehensible.**

Keywords- Salary prediction system; Motivation; Decision Tree model; graduated student history

I. INTRODUCTION

Nowadays, most university students enter into academic life without proper goals so they lack motivation to study in a class. Therefore, most students get bored of studying, and this causes failure in understanding lesson and examination [1]. This problem makes students to miss a required competency and knowledge which causes them retired from university. Apparently, students start their university life because they expect to have a good career with high payment after their graduation. Since comfortable life with sufficient income is one of most common human dream, the idea of this research is to use a predicted salary from graduated student history to motivate the eagerness to study and work toward their future plan.

For students to observe a salary rate in Thailand, students normally set their baseline salary rate from job recruiting companies such as Adecco, the world's leading human recruiting company. Recruiting companies collect salaries from survey, salary guide, and market insight from the workforce. Though recruiting companies provide Thailand salary guides every year, the guides only display the demand specific of employers and represent salaries using tables and graphs with many insight details based on economical domain wordings. This representation becomes difficult to understand for common people. Moreover, the guides are for overall viewpoint and do not support an individual salary rate based on a personal profile [2] or commonness of the profiles.

In fact, the realistic way to give salary information to students is to suggest based on existing data of graduated students. There are several researches on this topic such as Young-joo Lee et al. [3]. They applied regression method to predict not only salary but also satisfaction in jobs based on graduated student's history data. Moreover, the generated regression models match students and jobs based on student educational history and job descriptions. John jerrim [4] proposed an ordinary least-squares (OLS) regression model to conduct prediction models of future salaries for American college students based on family background data, profile enrolls and so on. According to the results of their analysis, they suggest that the prediction models based on 20-year-old college men are more realistic for predicting future salary of young people. However, the issue of those works is that the results representing in tables containing many statistical variables which are difficult to users without statistical background to comprehend.

On the other hand, Karla et al. [5] presented their work by using the prediction model of salaries for students in Master of Business Administration (MBA) program. They use hierarchical linear regression to conduct the prediction model with student and program characteristics as control variables and salary as the predictor variable. However, this prediction model predicts salaries as a group-based, and it can hardly be applied for an individual student.

According to the above-mentioned researches [3][4][5], although the proposed models perform well on a prediction of salaries task, but the problems are 1) the prediction model predicts salaries for a group, not for individual student and 2) the results from the prediction models require a person who has background in statistic to comprehend.

Therefore, in this paper, we propose a salary prediction system to predict students' future salaries based on graduated student history. We apply data mining techniques to conduct the salary prediction model by representing in a decision tree. Moreover, the system is designed to provide a history of graduated students with top-ranked salaries that similarly matched to user's profile in the simple user-readable display to

prevent incomprehensibility. We expect that this information will motivate the students to improve their study behavior.

The aim of this paper is to apply decision tree model for creating a salary prediction system. We expect that the system will increase motivation of students to study. The rest of this paper is organized as follows. Related works are described in Section II. The proposed system is explained in details in Section III. Section IV shows experiment results including accuracy result and effect result. Finally, we conclude the paper in Section V.

II. LITERATURE REVIEW

In this paper, we motivate student to improve their study behavior using salary prediction.

To survey salary, [2] Adecco Group collects salaries from survey, salary guide, and market insight from the workforce and provides Thailand salary guides every year such office position, job position, job description more than 750 position and among tables salary 74 tables. Nevertheless, they only show the demand specific of employers and represent salaries using tables and graphs with many details so it is difficult to understand but they cannot predict salary for each student based on his/her profile.

Young-joo Lee et al. [3] apply regression method to predict salary and satisfaction in jobs based on graduated student's history data and job descriptions. Afterwards, John jerrim [4] apply an ordinary least-squares (OLS) regression model to conduct prediction models of future salaries for American college students based on family background data, profile enrolls and so on. There are more realistic for suggest young people in employment actually make predictions of their future salary. The sample data is 4,434 data from collected data.

Then, Karla et al. [5] purpose the prediction model of salaries for students in Master of Business Administration (MBA) program. They use history data of 976 students in MBA programs and graduate business schools over the years 1988-2005. They use hierarchical linear regression to conduct prediction model with student and program characteristics as control variables and salary as the predictor variable. Though this prediction model predicts salaries for a group, it does not a prediction model for individual student.

From the above-mentioned papers, they perform the performance evaluations and the results show that the proposed models perform well prediction of salaries. However, the remaining problems are as followed:

1.) The prediction model predicts salaries for a group, not for individual student and

2.) The results from the prediction models are represented using tables containing statistical variables so the users without statistical knowledge are difficult to understand the results.

To solve the above problems, this paper proposes a salary prediction system that predicts salary using data mining techniques and shows predicted result in easy way to understand without requiring statistical knowledge.

III. SALARY PREDICTION SYSTEM USING DECISION TREE

In this paper, we propose a salary prediction system. We survey data mining techniques that are used in educational researches.

Maomi Ueno [6][7] proposes a pedagogical agent based on decision tree model constructed from the learning history data in the database to predict a student's future final status (1. Failed, 2.Abandon, 3.Successful and 4.Excellent). He demonstrates the effectiveness of this system through actual student data. The results show that the average score of the final test in the case of the developed e-learning system with the agent is significantly higher than the average score of the final test in the case of the e-learning system without the agent.

Farhana Sarker et al. [8] introduces decision tree model to predict students mark based on institutional internal and external open data sources that can be used in practical settings to predict students' academic performance. The result of this study shows that the prediction model based on institutional internal databases and external open data sources performs better than the prediction model based on only institutional internal databases.

Dorina Kabakchieva [9] made a study on comparing data mining techniques on data of education domain. The aim of the work was to find a technique that performs best on suggesting suitable campaign decisions for groups of different students from data collected by universities. The achieved results reveal that the decision tree classifier outperformed other due to the highest accuracy for prediction from all classes.

Zlatko J. Kovacic [10] identifies the important factors for student success and profiles of the typical successful and unsuccessful students. Moreover, Kovacic constructs prediction models using decision tree based on the important factors and student history data. The system evaluation results show that the prediction model provide high prediction accuracy.

The previous researches indicate decision tree models are popular data mining techniques to be applied as prediction models in educational research area. Moreover, the prediction models using decision tree models are easy to understand without statistical knowledge.

Therefore, we apply a decision tree model to conduct salary prediction model based on graduated student history data. In this paper, we use the history data collected for 10 years since 2006-2015 from Rajamangala University of Technology Thanyaburi, Thailand.

The history data include gender, faculty student enrolled, program student enrolled, type of their work-places, experience of job training, received certification, total Grade Point Average (GPA), salary, address, telephone number, e-mail etc.. To connect the data, the history data are represented in a form of decision tree.

In general, a decision tree consists of three components: 1) leaf nodes which represents decision output, 2) internal nodes, and 3) decision branches (lines) [11][12]. We determine salaries as our final output so we assign them as leaf nodes

and other data as attributes so they are put as internal nodes. The history data contain 22,540 rows. These data are used for training for salary prediction model.

Unfortunately, the history data have accumulated for 10 years of entries, and the base salaries have relatively increased each year based on an economic value. Therefore, the former salaries cannot be directly used to create salary prediction model for recent students. Therefore, we apply linear equating techniques to adjust the previous salaries based on current base salary where our hypothesis is that the distributions of salaries are similar for every year. By step-wise, data preparation process is firstly needed to be conducted. Afterwards, the salary prediction model will be created and used in the salary prediction system.

A. Data Preparation

The history data apparently contain several missing values and excessive amount of attributes for creating salary prediction model. Moreover, the data format is not proper for data mining tool. Therefore, we prepare the data as three following steps:

1.) Data Selection: We select relevant attributes for salary prediction and we use forward selection [13] analysis to select 7 attributes from 108 attributes. This greatly reduces data complexity and increases speed of processing. The remaining 7 attributes are as follows: *gender, faculty, program, type of works, job training, certificate,* and *GPA.*

2.) Data Cleaning: Outlier data (e.g. noticeably high salary) and missing values (e.g. no salary records) are manually removed. After data cleaning, the remaining history data are 13,541 rows.

3.) Data Transformation: This step prepares the data format to be usable for data mining tool. The data of salaries and GPAs are changed from continuous to interval format. We apply user specific discretization [14] of the salaries in which the salaries are divided to four levels for classes and the GPAs are divided into two levels. There are four levels of salary, less than 13,500, 13,501 - 15,300, 15,301 - 18,000 and More than 18,000. There are two levels of GPA which are < 2.8 and > 2.8.

B. Model Creation

In this paper, we create a decision tree model for predicting salary using data mining tool called Rapid Miner Studio 7 [15]. The example of decision tree model is shown in Fig. 1. The root node is Gender and has two branches that are left node as Male and right node as Female. Once Gender is focused on *Female*, the next node becomes node of *Type of works*. Once Type of works is focused on Employ of company, the next node becomes node of Faculty. Once Faculty is focused on Engineering, the next node becomes node of Program. Once Program is focused on Engineering Civil. The next node becomes leaf node of *> 18,000.*

Since, the decision tree model cannot be directly installed in the salary prediction system. We change the decision tree of salary prediction model to rules. The result is 810 rules. For example, the decision tree model in Figure 1 is changed to the rule as followed: if Gender = Female and Type of Works =

Employ of company and Faculty = Engineering and Program = Engineering Civil then salary is >18000. These rules are applied for creating salary prediction system in next step.

C. Salary Prediction System

The goal of this system is motivation student using 1) the salary prediction from profile of students, 2) the examples of three graduated students who have similar profile and the top-ranked salaries. Accordingly, users should understand the predicted results without statistical background.
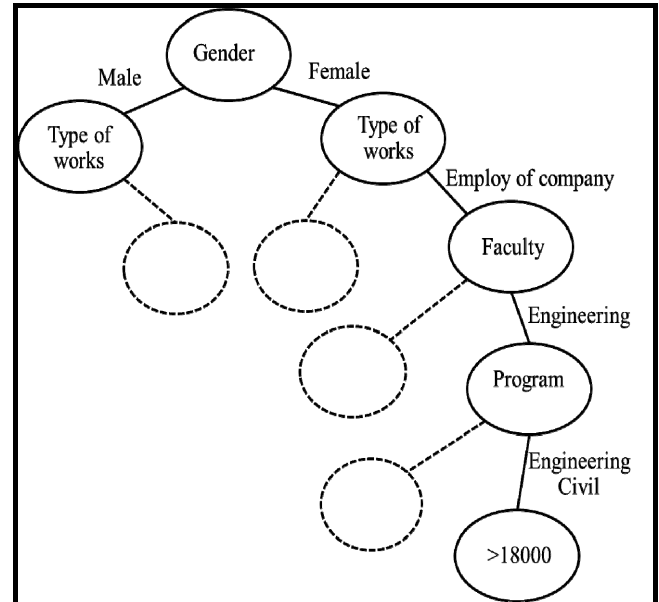

Figure 1.   Result of Decision Tree

Figure 2. Interface of Salary Prediction System

Fig. 2 illustrates the interface of Salary Prediction System. The interface contains seven attributes mentioned before. In

this example, they are gender, faculty, program, type of works, job training, certificate, GPA. These inputs are considered as the attributes that require users to fill in the salary prediction system. After the users input the attributes and click on submit button on bottom of the interface, the system compare the inputs of attributes with the rules and show predicted salary on the top of result interface as shown in Figure 3. For example in Fig. 3, the predicted salary is *more than 18,000 baht*. Moreover, the salary prediction model compares the inputs of attributes with the profile in the database and selects the three graduated students with similar attributes with the highest salaries. The system shows profiles of the three graduated students as follows: Gender, Faculty, Program, Type of work, job training, Certificate, GPA, Skill of ability and salary.

Figure 3. Salary Prediction Result

Therefore, we can conclude that the proposed salary prediction system has the potential to solve the problems of the previous systems as follows.

1) The salary prediction system can predict salaries of individual student.
2) The users do not require statistical knowledge to understand the results from the salary prediction system.

IV. EVALUATION AND RESULT

In this section, content is divided into three parts. The first part provides an evaluation of the salary prediction model by precision and recall measurement. The second part is an evaluation to measure the change of student motivation based on questionnaires. The third part is a satisfaction score from users upon the use of the salary prediction system.

A. Evaluation of the salary prediction model.

We use 10-fold cross-validation technique to evaluate the efficiency of the salary prediction model created using data described in Section III. Table 1 indicates the evaluation results of the salary prediction model in terms of F-measure, Precision, Recall, and Overall Accuracy [16][17].

TABLE I. SUMMARY OF SALARY PREDICTION MODEL

Class	F-measure (%)	Precision (%)	Recall (%)	Overall Accuracy (%)
Less than 13,500	46.22	42.97	50.01	41.39
13,501 - 15,300	36.03	36.03	36.03	
15,301 - 18,000	31.82	36.07	28.32	
More than 18,000	**50.52**	**49.15**	**51.97**	

According to the result shown in Table I, the best results yielded from the system are on a class of salary *More than 18,000* as 50.52, 49.15 and 51.97 points for f-measure, precision and recall respectively. For overall accuracy, we got 41.39%. In detail, we found that the features of class '*salary More than 18,000*' are more commonly similar within themselves comparing to the features of the other classes. This circumstance led to the high result in terms of Precision and Recall, and resulted in the highest F-measure among all classes.

B. Evaluation of student motivation

In this part, questionnaire was selected to evaluate student motivation. It is divided into three parts. The popular samplings are 50 students from Rajamangala University of Technology Thanyaburi, Thailand, which is the same source of those in use in salary prediction system.

Part 1: general student information

This part gathers basic student information such as gender, age, faculty and program. The obtained data are summarized as follows. For gender, we got 26 male and 24 female samplings. A range of age is 19 to 22 year-olds. From a faculty perspective, students are from 7 different faculties. For academic year, we got 7, 15, 5, and 23 student samplings from year 1 to 4 respectively.

Part 2: question to examine student motivation before using this part contains questions to examine student motivation. The questions and students' result are shown in Table II.

TABLE II. QUESTION BEFORE USING THE SALARY PREDICTION SYSTEM AND RESULTS

Question No.	Question	Scores / Answers	Mean(SD)/Full scores
Q1	Does your current faculty/program match your objectives?	1.Yes, 2.No	1.26 (0.44) / 2
Q2	Which level of satisfaction do you have for your faculty/ program?	1. Least, 2. Little, 3.Moderate, 4.Much, 5.Most	3.8 (0.57) / 5
Q3	Does the high salary affect your study motivation?	1. Least, 2. Little, 3.Moderate, 4.Much, 5.Most	4.16 (0.47) / 5
Q4	Which salary rate do you expect after graduation?	1.Least than 13,500 2. 13,501 - 15,300 3. 15,301 - 18,000 4. More than 18,000	3.7 (0.34) / 4

Part 3: question to examine student motivation after using

This part contains questions to examine student motivation after using the system. The questions and students' results are shown in Table III.

TABLE III. QUESTIONS AFTER USING THE SALARY PREDICTION SYSTEM AND RESULTS

Question No.	Question	Scores / Answers	Mean(SD)/ Full scores
Q5	How much does the predicted salary from the system match your expected salary?	1. Least, 2. Little, 3.Moderate, 4.Much, 5.Most	4.00 (0.45) / 5
Q6	How much satisfaction do you have with your current faculty/program?	1. Least, 2. Little, 3.Moderate, 4.Much, 5.Most	4.16 (0.55) / 5
Q7	After knowing the predicted salary, how much satisfaction do you have with your enrolled faculty/program?	1. Least, 2. Little, 3.Moderate, 4.Much, 5.Most	4.12 (0.59) / 5
Q8	How much does the system motivate you to study for your expected salary?	1. Least, 2. Little, 3.Moderate, 4.Much, 5.Most	4.18 (0.40) / 5

From the Q1, the majority of answers is that most of students are on the track of their objectives in their current faculty and program. Q2's mean score results show that students are satisfy in high degree for their current faculty and program. Moreover, in comparison Q2 with Q6, we found that the predicted salary can boost up students' satisfaction for their current faculty and program. From Q7's results, the predicted salary by the system makes students to realize the implicit benefit of being in the current faculty and program and an expedition to accomplish their own future plan. The Q8's results can be inferred that with the predicted salary, students have the much clearer goal and will likely be more motivated to study to achieve the expected salary.

C. Evaluation of salary prediction system

We evaluate the salary prediction system using questionnaires. The questionnaires contain the detail as shown in Table IV.

TABLE IV. AFTER USING THE SALARY PREDICTION SYSTEM

Question No.	Question	Scores / Answers	Mean(SD)/ Full scores
Q9	Is the system easy to use?	1. Least, 2. Little, 3.Moderate, 4.Much, 5.Most	4.02 (0.40) / 5
Q10	Is the system's result reliable in your opinion?	1. Least, 2. Little, 3.Moderate, 4.Much, 5.Most	3.89 (0.38) / 5

In this part, we mainly asked for an evaluation of satisfaction in using the system. By the given score, we found that the users are much satisfactory with the designed user interface as it is easy to use. For the reliability of the output, the yielded score is distinctively above average but not much trustable. This result is understandable since the prediction is for further ahead of their current state and very hard to be proved.

Moreover, we also asked for other suggestions related to the system as Q11. We can summarize the suggestions into three points as following:

- There are missing features and these should be added to improve the prediction. For examples, students suggested that co-operative and job-training should be split into two different features since it will really affect different salary rates.
- Some students suggested that this system should be used before they made an entrance examination so they can choose their academic path and gold rationally before making decision on faculty and program.
- Some students clearly mentioned that the prediction will make them more active and motivated in studying as they can noticeably see their future. Moreover, some students mentioned that results from graduated students can be an exemplar to them.

V. SUMMARY

In this paper, we proposed a salary prediction system for increasing students' motivation in studying. Decision tree technique was used to generate mining model for prediction with seven features. Moreover, the result of the system is not only a predicted salary, but also the 3-highest salary of the graduated students which share common attributes to the users. The results of the system were proved to be easily comprehensible for those students who do not have a statistical background. To test the system's efficiency, we set up an experiment by using 13,541 records of actual graduated student data in 10-fold cross validation. The total result in accuracy is 41.39%. Furthermore, we exploited questionnaires to evaluate effectiveness of the system with 50 student samples. The result shows that this system can boost students' motivation in studying and also show them a positive viewpoint of their future. At last, the sampling students answer the satisfaction score in positive manner since they found the

system was easy to use and the prediction result was simple and comprehensible.

To improve the system, we plan to include more features suggested by users or academic staffs and test them for effectiveness and validity. More training data will be collected to increase the coverage of occupations to other studying fields. Lastly, we plan to set up this service in a university site as consultant system for university entrance examinees.

REFERENCES

[1] Lumsden, Linda S., "Student Motivation To Learn. ERIC Digest", Number 92., 1994.

[2] Adeco Office in Thailand , Adecco Thailand Salary Guide, 2015.

[3] Young-joo Lee & Meghna Sabharwal , " Education–Job Match, Salary, and Job Satisfaction Across the Public,, Non-Profit, and For-Profit Sectors: Survey of recent college graduates", Public Management Review, 18:1, 40-64., 2014.

[4] John Jerrim, "Do college students make better predictions of their future income than young adults in the labor force?", Education Economics, 23:2, 162-179. 2013.

[5] K. R. Hamlen,W. A. Hamlen,"Faculty salary as a predictor of student outgoing salaries from MBA programs", Journal of Education for Business, 91:1, 38-44. 2015.

[6] M. Ueno, "Animated Pedagogical Agent based on Decision Tree for e-Learning", Proc.IEEE conference (Computer Science), ICALT ,2005.

[7] M. Ueno, "Animated agent to maintain learner's attention in e-learning" , Proc. E-Learn, 2004.

[8] F. Sarker, T. Tiropanis, H. C. Davis, "Students' Performance Prediction by Using Institutional Internal and External Open Data Sources", CSEDU, 2013. page 639-646. SciTePress.

[9] D. Kabakchieva, "Predicting Student Performance by Using Data Mining Methods for Classification", CYBERNETICS AND INFORMATION TECHNOLOGIES, Volume 13, No 1., 2013

[10] Z. J. Kovacic,"Early Prediction of Student Success: Mining Students Enrolment Data", Proceedings of Informing Science & IT Education Conference (InSITE), 2010.

[11] J.R. Quinlan, "Induction of Decision Trees", 1986 Kluwer Academic Publishers, Boston, Manufactured in The Netherlands, Machine Learning 1:81-106, 1986.

[12] T. Theeramunkong,"Introduction to Concepts and Techniques in Data Mining and Application to Text Mining", Second Edition, 2012.

[13] M. A. Hall, G. Holmes."Benchmarking Attribute Selection Techniques for Discrete Class Data Mining", IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL.15, NO. 3, 2003.

[14] P. Berka, I. Bruha,"Discretization and grouping: preprocessing steps for Data Mining", Principles of Data Mining and Knowledge Discovery, Second European Symposium, PKDD '98, Nantes, France, 1998.

[15] Eakasit Pacharawongsakda,"Practical Data Mining with RapidMiner Studio 7",#12,Tutorial documentation, 2016.

[16] O. Villacampa,"Feature Selection and Classification Methods for Decision Making: A Comparative Analysis", A dissertation submitted in partial fulfillment of the requirements for the Degree of Doctor of Philosophy In Information Systems, College of Engineering and ComputingNova Southeastern University , 2015.

[17] G. Forman,"An Extensive Empirical Study of Feature Selection Metrics for Text Classification", Journal of Machine Learning Research 3, 1289-1305, 2003.