Token Allocation for Course Bidding With Machine Learning Method

Chonbadee Juthamanee
Department of Computer Engineering
Chulalongkorn University
Bangkok, Thailand
6270047421@student.chula.ac.th

Krerk Piromsopa
Department of Computer Engineering
Chulalongkorn University
Bangkok, Thailand
krerk@cp.eng.chula.ac.th

Prabhas Chongstitvatana
Department of Computer Engineering
Chulalongkorn University
Bangkok, Thailand
prabhas.c@chula.ac.th

***Abstract*—Most educational institutions have a problem of allocation course seats to students because demand exceeds supply for many courses. A course bidding system is a tool for improving the registration system based on auction theory to provide course seat allocation equitably and efficiently. This paper proposes a method for token price prediction for allocate to courses through course bidding system. Three methods are compared: Decision Tree, Random Forest and Artificial Neuron Network. The result of the experiment shows that ANN is the best method with lowest RSME 3.98%. Furthermore, it provides an important information to users to estimate the risks on their course bidding strategy.**

Keywords— Course Bidding, Bid Allocation, Decision Tree, Random Forest, Artificial Neuron Network

I. INTRODUCTION

Auction Theory is an economic science for allocating limited resources in the market between seller and buyer for the best effectiveness and performance, such as some easy items auction, who give the highest price get an item, while some procurement in the government auction will choose who gives the lowest price, and so on. [2] Each market will be different and varies by the rule base on three compositions, 1) Auction type 2) Items type 3) an Information that everyone can access and uncertainly. An auction is a tool used in many industries. Many players try to find the solution to win the auction game.

Jie Mei, Dawei He, Ronald Harley, and Thomas Habetler [3] proposed Random Forest (RF) and ANN methods to forecast real-time prices in the New York electricity market to win power bids in highly competitive market. The method provides a price probability distribution, allowing users to estimate the risks of their bidding strategy and making the results helpful for later industrial use. The result from RF shows the best error at 12.03%

Jong-Min Kim and Hojin Jung [4] proposed regularized linear regression method to estimate interval for wining project on highway procurement auctions. They are using random forest to select the important tasks and used them for input. The result is compared with the least square linear model based on the bias, and the standard root mean square error of the bid estimates. It shows that the suggested approach provides superior forecasts for an interval of winning bids over the competing model.

Richard D. Lawrence [5] proposed a naive Bayes classification model to predict the bid outcome, win or loss, for optimal bid pricing by comparing with human pricing experts. They used an entropy-based information gain metric to extract features for predicting win/loss labels. The results show that a naive Bayes classification model has high win probability and is best to use for the optimal price. It is able to improve performance by getting information from human pricing experts.

In education section, some institutions used auction theory to apply to the course register system for allocating course seats equitably and efficiently while demand exceeds supply for many courses [6].

However, there are few tools and research about winning the game, as examples above. Therefore, we propose a method to predict the price for allocating to the course to win and to plan the strategy by using machine learning. We compare three methods, Decision tree, Random Forest, and ANN.

II. TOKEN ALLOCATION FOR COURSE BIDDING

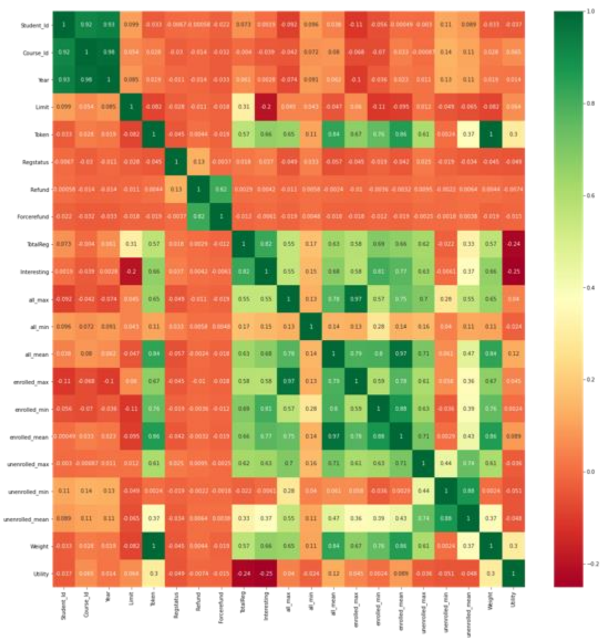
In this section, we use data from the course bidding system with each student is given a million tokens as bid endowment to allocate the courses he considers taking. Every semester there are 225 students and 28 subjects on average.

The process of prediction consists of main three steps as follows:

a. Pre-Processing. The size of the data is a total of 5,398 transactions. We selected only winner bidders and calculated course interesting values with total registered student divide with total seats, and calculate important statistic values as max, min, and mean on the winning bidders and losing bidders. This statistic improves an overall transaction on every subject to give more information to the model under normalize form. Figure 1 shows the correlation metric of all features. The features that are strongly related with a score of more than 0.60 and have at least 4 relations on each other features are selected to use for training. Table 1 shows 8 input variables and 1 output variable to predict is a token price used in training the model. The value of 'course interesting' is real number. All other input values are normalized.

Table 1. Input and output variable

input	output
'course interesting', 'all_max', 'all_mean', 'enrolled_max', 'enrolled_min', 'enrolled_mean', 'unenrolled_max', 'unenrolled_mean'	'Token Price'



High (green) = 1.0 / Low (red) = 0.0

Figure 1. Correlation Metric of all variables

b. Training Model. We compare 3 models decision tree, Random Forest (RF) and ANN that was experimented by many [3, 11, 12]

Decision Tree is popular and commonly used in ML for establishing classification and performs well on regression problems. It classifies a population into branch-like segments and constructs an inverted tree with a root node, internal nodes, and leaf nodes [7].

Figure 2 displays RMSE with the depth of trees from 1 to 20 and set other parameters to default values.

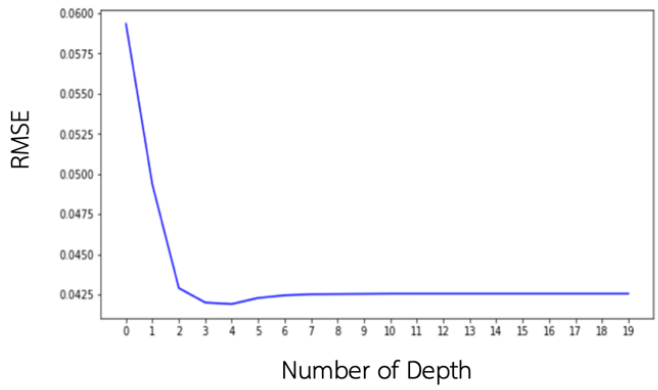


Figure 2. RMSE on Decision Tree

Random Forest is a popular machine learning method. This method is an ensemble of classification algorithm, which uses trees as base classifiers. It performs well on classification and regression with many different types of datasets [8]. RF is based on the bootstrap and selection of random subset of predictor variables as candidates for splitting tree nodes. This method constructs many decision trees with many individual learners combined. It is also known that RF produces good out-of-sample fits for highly nonlinear data [9].

Figure 3 displays RMSE with the number of trees obtained by RF. We make a RF of 100 trees and set depth of tree at 4 to minimize prediction error on evaluation.

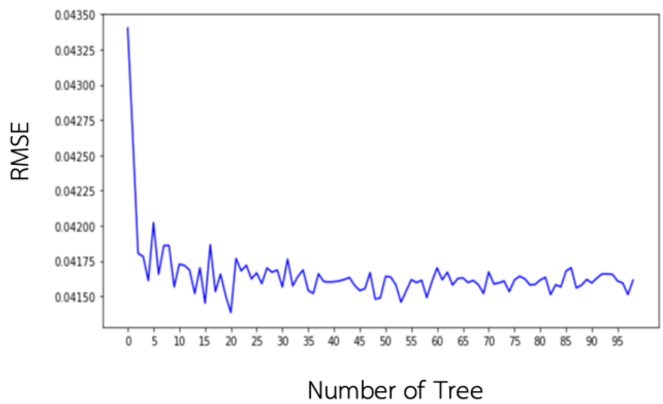


Figure 3. RMSE on Random Forest

ANN is designed to acquire knowledge by extracting useful patterns from data. It is used in many sectors—economics, forensics and pattern recognition [10].

We set a vector including 8 variables as input. Each neuron in the first node layer is connected to each of the elements in an input vector. There are 2 hidden layers with Relu activation function, 8 neurons in the first hidden layer, 8 in the second layer. The output at the last node is a token price prediction on each course.

Figure 4 displays RMSE with 100 epochs during training and validation with real value and prediction value every transaction ordered in a single list. An average error with 10-fold validation is 4.14%

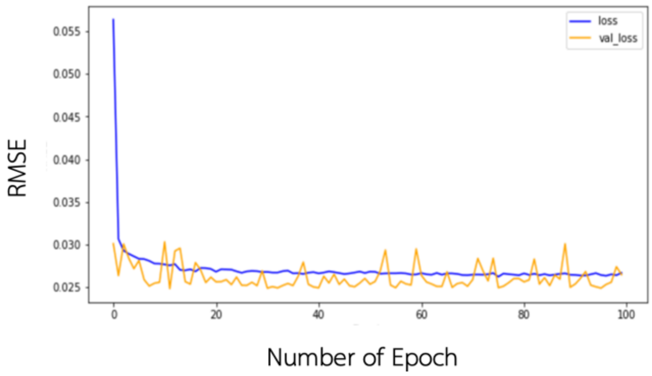


Figure 4. RMSE on ANN

c. Evaluation

We evaluation each method with test sets using the best model that was already trained by training set and choose from the lowest RMSE, then compare prediction values with real values.

III. EXPERIMENT RESULT

The experimental results are shown in Table 2. The table shows error with RMSE on the test set. Decision Tree has highest error at 4.18%, Random Forest error is lowest than decision tree 4.13%, while ANN shows lowest error at 3.98%. ANN is the best performance method to predict token for allocation in course bidding.

Next, we use this result to design a tokens recommendation system and share it on public and allow every user to plan the strategy. For example, if a user wants to register for the most popular course, this system will suggest which related course suitable for the schedule along education semester.

Table 2. Experiment Result

Method	RMSE
Decision Tree	4.18 %
Random Forest	4.13 %
ANN	3.98 %

IV. CONCLUSION

This study presents a method for token price prediction for allocating to courses through a course bidding system. We perform experiments with real data with three different machine learning methods for course bidding. The result shows ANN is the best method with the lowest RSME 3.98%, which mean that one has a probability of winning 96.02%.

[13] Feed-forward back-propagation in ANN has weight each input variable by using the connection weights algorithm (CW) to calculates the sum of products. This approach provides final weights that relative important to input variables.

While input weight calculating in decision tree and RF is unstable and significant changes on the models learning in sample values. Also, final tree in RF did not guarantee to be the optimal tree as well. [12]

REFERENCES

[1] Aradhna Krishna and M. Utku Ünver, “Improving the Efficiency of Course Bidding at Business Schools: Field and Laboratory Studies”, Marketing Science Vol. 27, No. 2, 2008, pp. 262-282

[2] Paul R. Milgrom and Robert B. Wilson, “Improvements to auction theory and inventions of new auction formats”, Scientific Background on the Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel, 2020

[3] J. Mei, D. He, R. Harley, T. Habetler and G. Qu, "A random forest method for real-time price forecasting in New York electricity market," IEEE PES General Meeting Conference & Exposition, National Harbor, 2014

[4] Jong-Min Kim & Hojin Jung, “Predicting bid prices by using machine learning methods”, Applied Economics Volume 51, 2019

[5] Richard D. Lawrence, “A Machine-Learning Approach to Optimal Bid Pricing”, Computational Modeling and Problem Solving in the Networked World, pp. 97-118

[6] Tayfun Sönmez & M. Utku Ünever. “Course Bidding at Business Schools”, International Economic Review Vol. 51, No. 1 2010, pp. 99-123

[7] Yan-Yan Song and Ying Lu, “Decision tree methods: applications for classification and prediction”, Shanghai Arch Psychiatry, 2015

[8] Leo Breiman, “Random Forests”, Statistics Department, University of California, Berkeley, CA, Machine Learning, 45, 5–32, 2001

[9] Hal R. Varian, “Big Data: New Tricks for Econometrics”, Journal of Economic Perspectives 28(2), 2014

[10] Larry Hardesty, “Explained: Neural networks”, MIT News, 2017
[11] Muhammad Waseem Ahmad, Monjur Mourshed, Yacine Rezgui, “Trees vs Neurons: Comparison between random forest and ANN for high-resolution prediction of building energy consumption”, Energy and Buildings, Volume 147, 2017, pp. 77-89
[12] E.J. Olaya-Marín, F. Martínez-Capel and P. Vezza, “A comparison of Artificial Neural Networks and Random Forests to predict native fish species richness in Mediterranean rivers”, Knowledge and Management of Aquatic Ecosystems, 2012
[13] O.M. Ibrahim, “A comparison of methods for assessing the relative importance of input variables in artificial neural networks”, Journal of Applied Sciences Research, 9(11): 5692-5700, 2013