

# Tree-based Machine Learning and Deep Learning in Predicting Investor Intention to Public Private Partnership

Ahmad Amin<sup>1</sup>, Rahmawaty<sup>2</sup>, Maya Febrianty Lautania<sup>3</sup>, Suraya Masrom<sup>4</sup>, Rahayu Abdul Rahman<sup>5\*</sup>

Faculty of Economics and Business-Universitas Gadjah Mada, Yogyakarta, Indonesia<sup>1</sup>

Faculty of Economics and Business-Universitas Syiah Kuala, Aceh, Indonesia<sup>2,3</sup>

Computing Sciences Studies-College of Computing-Informatics and Media, Universiti Teknologi MARA Perak Branch, Malaysia<sup>4</sup>

Faculty of Accountancy, Universiti Teknologi MARA, Perak Branch, Malaysia<sup>5</sup>

**Abstract**—Public private partnership (PPP) is the government initiate in accelerating public infrastructure development growth. However, the scheme exposes private sector to various risks including political risk which in turn affect financial performance and reporting of participating firms. Given that one of the issues facing the government is the lack of participation from the private sector in such arrangements. Thus, the main objective of this study is to observe the machine learning prediction models on private investor intention in participating the PPP program. Tree-based machine learning and deep learning are two different types of promising algorithms, which proven to be useful in widely domain of prediction problems but never been tested on the concerned problem of this study. Based on real data of investors for Indonesian listed firms, this paper presents the ability of the selected machine learning algorithms by means of different assessments point of view. First assessment is on the algorithms' performances in producing accurate prediction. Second assessment is to identify the variance of PPP attributes in each of the prediction model with the machine learning algorithms. The performance results show that all the prediction models with the machine learning algorithms and the PPP attributes were well-fitted at R squared above 80%. The findings contribute a significant knowledge to various fields of scholars to implement a more in-depth analysis on the machine learning methods and investors' prediction.

**Keywords**—Tree-based machine learning; deep learning; prediction; investor intention; public private partnership

## I. INTRODUCTION

The degree of development of a nation is determined by its capacity to provide for the needs of the general population, including infrastructure like power, internet, trains, roads, ports, and airports. Infrastructure investments may improve livelihoods and well-being since they make people's lives easier and better [1], [2]. However, many countries have several issues and difficulties when constructing public infrastructure. Financial factors, regional factors, demographic concerns, environmental difficulties, and human factors are some of the key causes of infrastructure development delays [3], [4].

In response to the financial constraint, the government has

initiated alternative funding programmes for the construction of public infrastructure and invited the private sector to participate. This special arrangement is known as public-private partnerships (PPP). PPP has developed in a number of developing countries since it was first used in the United Kingdom in the early 1990s, including Chile in Latin America, Colombia, Brazil, Bulgaria, and Slovenia in Eastern Europe, and Indonesia, Malaysia, and South Korea in Asia [5]–[7]. However, the degree to which PPPs are successfully implemented varies from country to country. For example, in Indonesia, despite the surge in PPP projects for infrastructure development, the private sector's participation is still low as the scheme dominated by Indonesian state-owned enterprise. Hence, this study aims to develop machine learning prediction model on private investor intention in participating the PPP program. This study has important contributions by extending prior works on PPP with non-financial factors to be analyzed based on the machine learning prediction models.

The following section provides a brief description on the data set of this study and the empirical steps of the machine learning research. Section III presents the results of the machine learning performances comparisons based on the PPP attributes. Finally, Section IV discusses the conclusions and future research directions.

## II. LITERATURE REVIEW

Prior research that applied machine learning on prediction, classification and detection studies in financial, accounting and finance research highlight the effectiveness and accuracy of such methods to that of traditional statistical methods in problems such as in detection of financial fraud [8], firm performance [9] and finance [10], [11]. Despite the wider use of machine learning in accounting and finance, yet study on machine learning prediction and classification on PPP investment is limited. To date, most of prior studies used machine learning to predict successful PPP projects [12]–[16] that highlighted the return benefits of deploying the intelligent approach for solving various issues of PPP. In [16], Random Forest has been identified as the most outperformed algorithm but the study has observed different aspects of PPP attributes that used in this study. Most of the researchers suggested that more advance research on the machine learning algorithms for

This research was funded by Ministry of Higher Education Malaysia under FRGS grant 600-IRMI/FRGS 5/3 (208/2019).

PPP is needed to accelerate the implementation of computational system for facilitating PPP projects and investments.

Prior studies on the determinants of PPP highlight several financial, economic and environmental factors such as general and operational risk [17],[18], return [18]; operational and cost recovery [19], ownership [20], risk of political instability [21]; risk of exogenous uncertainty [22] and environmental, development impacts and sustainability [20]. However, there are limited studies that have been conducted to examine how non-financial factors such as investor trust in the government, influence the decision of investors to finance public infrastructure through PPP arrangements. To provide a deepen current understanding on the private investor intention in participating in PPP programs, this study attempted to discover non-financial factors including government trust, service quality, transparency, and values of similarity as the determinants of PPP. Unlike the prior works, this paper presents the findings of analysis from the implementation of machine learning prediction models.

To date, more than hundreds of machine learning algorithms can be utilized for various domains of problem such as Decision Tree [23], Random Forest [24], Support Vector Machine [24] and Gradient Decision Trees [23]. Decision Tree, Random Forest and Gradient Decision Trees are categorized as tree-based machine learning. The tree-based machine learning is robust to be used for regression and classification problems. Additionally, Deep Learning [25] is very promising algorithm for prediction problems with massive attributes. However, to study its ability on simple attributes as proposed in this study will provide another useful knowledge to researchers from various fields of interest. The findings will be beneficial to the relevant PPP stakeholders for implementing vast and rapid data-driven recommendations and decisions.

### III. METHODS

#### A. Dataset

The machine learning algorithms were tested on dataset that consists of data from 165 top management of Indonesian listed firms. Based on the real collected data, Pearson correlation test was conducted to observe the weights of correlation coefficient each of the PPP attributes as the independent variables (IVs). Fig. 1 presents the PPP attributes correlation coefficient to the dependent variable (DV). The DV in the predictor model is the investor intention to the PPP. The PPP attributes is the investment intention factors namely based on government trust (*TrustOnGov*), perceived of government service quality (*GovServiceQ*), perceived of government transparency (*GovTransparency*), and similarity of values (*ValueSimilarity*).

Two of the PPP attributes present positive strong correlations (above 0.7 correlation coefficient) to the investor intention while the rest of two attributes have moderate correlation coefficient (0.5-0.6). In the machine learning predictive models, one important matter that need to be observed is the contribution of each attribute in providing

knowledge for the machine learning to make prediction. Two questions to be answered in this research are:

RQ1: Are all the four PPP attributes usable to each of the machine learning prediction model?

RQ2: Which of the attributes are mostly important to all the machine learning models?

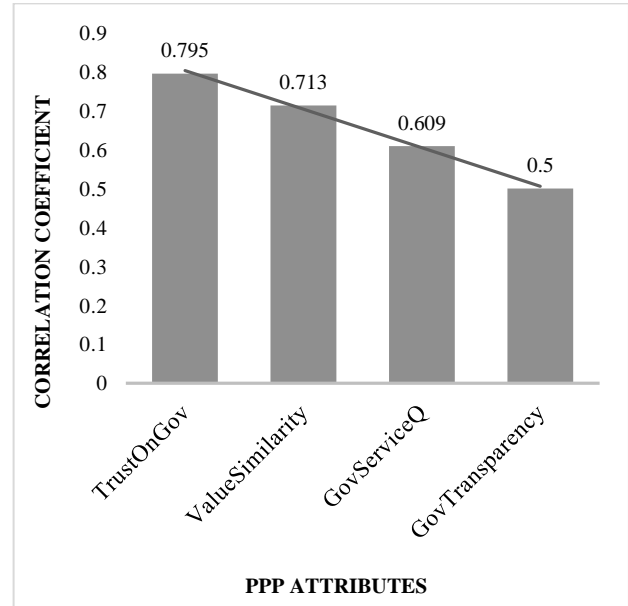


Fig. 1. Weights of correlation each PPP attributes to the DV.

#### B. Machine Learning

Two machine learning from the family of tree-based algorithms namely Random Forest (RF) and Decision Tree (DT) were used in this study. Additionally, acknowledging the wider used of Deep Learning algorithm, it is interesting to observe the algorithm ability in the case of PPP investor intention. Based on preliminary experimental series, the most optimal hyper-parameters of RF and DT is listed in Table I.

The range of number of trees for observing the RF accuracy were 20, 60, 100,140. For each combination, the worst error rate produced by RF was 11.6% while the best error rate was 10.5% with number of tress 140 and maximal depth 4. For DT, the range of maximal depth used in the preliminary testing is between 2 and 25. The most optimal value has been achieved by DT when the maximal depth was 4 at error rate 11.8%, slightly lower from the maximal depth value 2(11.9%). A similar value of error rate at 12.8% has been produced when the maximal depth was individually set as 7, 10,15 or 25. The production of prediction generated by Deep learning is depicted in Fig. 2. The number of layers is 4 that used Rectifier function for the input layers and Linear function as the output layer (layer 4).

TABLE I. THE OPTIMAL HYPER-PARAMETERS

Machine Learning Algorithm	Hyper-parameters	Error rate %
RF	Number of trees=140 Maximal depth=4	10.5
DT	Maximal depth=4	11.8

```

root mean squared log error: 0.106737055
Status of Neuron Layers (predicting InvestmentIntention, regression, gaussian distribution, Quadratic loss, 2,851 weights/biases, 38.3 KB,
Layer Units      Type Dropout      L1      L2 Mean Rate Rate RMS Momentum Mean Weight Weight RMS Mean Bias Bias RMS
1      4      Input  0.00 %
2      50 Rectifier  0 0.000010 0.000000 0.002733 0.001804 0.000000 -0.003671 0.187752 0.496365 0.013304
3      50 Rectifier  0 0.000010 0.000000 0.007516 0.014828 0.000000 -0.001072 0.138856 0.997919 0.013785
4      1      Linear  0.000010 0.000000 0.000333 0.000157 0.000000 0.024648 0.204335 -0.007682 0.000000

Scoring History:
Timestamp      Duration Training Speed      Epochs Iterations      Samples Training RMSE Training Deviance Training MAE Training r2
2022-11-02 10:33:25 0.000 sec      0.000000      0 0.000000      NaN      NaN      NaN      NaN
2022-11-02 10:33:25 0.013 sec      20625 obs/sec 1.000000      1 165.000000 0.49599      0.24600      0.37078      0.56758
2022-11-02 10:33:25 0.024 sec      19411 obs/sec 2.000000      2 330.000000 0.53033      0.28125      0.40805      0.50562
2022-11-02 10:33:25 0.034 sec      19038 obs/sec 3.000000      3 495.000000 0.45317      0.20537      0.34044      0.63901
2022-11-02 10:33:25 0.045 sec      18857 obs/sec 4.000000      4 660.000000 0.46653      0.21765      0.34810      0.61742
2022-11-02 10:33:25 0.055 sec      19186 obs/sec 5.000000      5 825.000000 0.47776      0.22826      0.36527      0.59877
2022-11-02 10:33:25 0.065 sec      19411 obs/sec 6.000000      6 990.000000 0.44285      0.19612      0.32818      0.65527
2022-11-02 10:33:25 0.076 sec      19250 obs/sec 7.000000      7 1155.000000 0.44362      0.19680      0.32992      0.65407
2022-11-02 10:33:25 0.086 sec      19411 obs/sec 8.000000      8 1320.000000 0.43911      0.19282      0.32633      0.66106
2022-11-02 10:33:25 0.098 sec      19285 obs/sec 9.000000      9 1485.000000 0.44436      0.19745      0.33222      0.65292
2022-11-02 10:33:25 0.112 sec      18333 obs/sec 10.000000     10 1650.000000 0.44296      0.19621      0.33480      0.65511
2022-11-02 10:33:25 0.117 sec      17741 obs/sec 10.000000     10 1650.000000 0.43911      0.19282      0.32633      0.66106
    
```

Fig. 2. The production of deep learning.

#### IV. RESULTS AND DISCUSSION

Table II presents the performances comparison between the three machine learning algorithms by means of Root Mean Square Error (RMSE), R squared ( $R^2$ ) and the time required to complete (TTC) the processes of training and predicting. RMSE measures the accuracy error of the machine learning models while  $R^2$  is presenting the proportion of IVs variation in the prediction models.

Table II shows that all the predictive models with the PPT attributes have contributed at least 80% of the fitted data from the  $R^2$ . Random Forest provided the best results in line with the results in [16] but on this cases that used different PPP attributes, the  $R^2$  is lower that the results in [16]. Although the average error presented by the Relative Error and RMSE of all models are not very encouraging, the values are within an acceptable range to anticipated that the machine learning can relatively predict the data accurately. All the algorithms are reliable when tested on the PPP dataset as described through the results of small standard deviation ranges.

In term of efficiency, DT has taken the shortest time to complete followed by Deep Learning and RF. Furthermore, it will be useful to look the variance of each PPT attributes in the machine learning predictive models as listed in Table III.

As seen in Table III all the PPT attributes were contributed some degree of correlations in the prediction models of all the machine learning algorithms. Thus, all the PPP attributes used in this study were used by the machine learning algorithms, than answered the RQ1. Deep Learning received bigger weight of contributions from the trust on government and value of similarity. For all algorithms, trust on government has given a major influence the prediction models at weight above 0.5, which answering the RQ2 of this research. The lowest influence from the PPP attributes can be seen from the government service of quality. Furthermore, the following Fig. 3 and Fig. 4 show the Tree models generated by DT and RF, respectively.

TABLE II. THE PERFORMANCES RESULTS

Algorithm	Relative Error (+-std.dev)	RMSE (+-std.dev)	$R^2$ (+-std.dev)	TTC (ms)
RF	9.9% (3.8%)	0.487 (0.108)	0.826 (0.094)	553
DT	10.3% (2.5%)	0.5 (0.108)	0.802 (0.114)	61
Deep Learning	10.4% (2.5%)	0.477 (0.098)	0.824 (0.056)	342

TABLE III. WEIGHTS OF CORRELATION BETWEEN PPP ATTRIBUTES IN MACHINE LEARNING

Algorithm	Trust on Government (TrustonGov)	Value of Similarity (ValueSimilarity)	Government Transparency (GovTransparency)	Government Service of Quality (GovServiceQ)
RF	0.674	0.395	0.078	0.053
DT	0.597	0.230	0.055	0.048
Deep learning	0.717	0.662	0.028	0.029

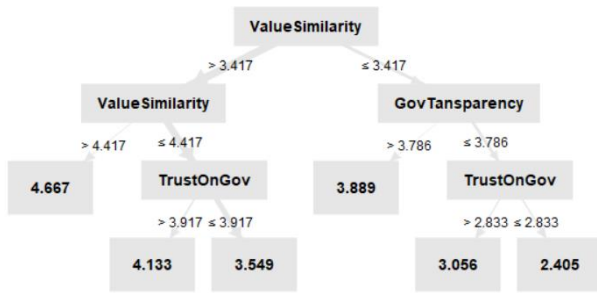


Fig. 3. The decision tree.

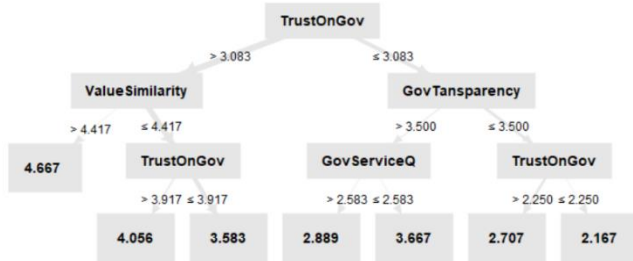


Fig. 4. The random forest tree at 140 numbers of trees.

The depth of the tree models from DT and RF is 4 as set at the optimal maximal depth. The trust on government is the deeper attribute before the leaf of both trees that representing it as the important contribution to the model. Although the weights of similarity of value and government transparency are very low in DT and RF, the two attributes are still included in the tree models. Otherwise, government service of quality is utilized in RF but not in DT.

## V. CONCLUSIONS

This study aims to explore the non-financial issues that have influenced the investor intention in the public-private partnerships (PPP) based on the evidence from Indonesia. Acknowledging the role of machine learning to provide fast data-driven prediction and how complex is the challenge of future data expanding, this study attempted to discover machine learning techniques and has observed the effect of the pre-determined non-financial factors in the successful of PPP including Government Trust, Similarity of Values, Transparency in Government and Service Quality of the Government. Within the scopes of the tested data based on PPP in Indonesia, the attribute of Government Trust presents the most significant factor in the prediction models inside and outside of the machine learning implementations. This study can be further extending in the future by considering more PPP attributes as suggested in [16] that can improve the accuracy of the machine learning models and use different machine learning algorithms than the proposed study in this paper.

## ACKNOWLEDGMENT

This research was supported by the Ministry of Higher Education Malaysia under FRGS grant 600-IRMI/FRGS 5/3 (208/2019). Thank you very much to the Ministry of Higher Education Malaysia and the Universiti Teknologi MARA for the full support.

## REFERENCES

- [1] H. Kociemska, "Public-Private Partnership: Reconciling Mainstream and Islamic Finance in sub-Saharan Africa," *Emerg. Mark. Financ. Trade*, vol. 56, no. 12, pp. 2891–2907, 2020.
- [2] S. Y. In, L. A. S. Casemiro, and J. Kim, "A decision framework for successful private participation in the airport sector," *J. Air Transp. Manag.*, vol. 62, pp. 217–225, 2017.
- [3] I. Walter, "The infrastructure finance challenge," Available SSRN 2841281, 2016.
- [4] S. S. Clark, T. P. Seager, and M. V Chester, "A capabilities approach to the prioritization of critical infrastructure," *Environ. Syst. Decis.*, vol. 38, no. 3, pp. 339–352, 2018.
- [5] S. Pfisterer, "Public-Private Partnership for Development: Governance Promises and Tensions," in *The Emerald Handbook of Public-Private Partnerships in Developing and Emerging Economies*, Emerald Publishing Limited, 2017.
- [6] R. Mohamad, S. Ismail, and J. M. Said, "Performance objectives of public private partnership implementation in Malaysia: perception of key players," *J. Asia Bus. Stud.*, 2018.
- [7] S. Ismail, M. S. Musawa, and H. Ahmad, "Transparency of public private partnership (PPP): the extent of mandatory information disclosure," *Built Environ. Proj. Asset Manag.*, 2019.
- [8] A. Priyadarshini, S. Mishra, D. P. Mishra, S. R. Salkuti, and R. Mohanty, "Fraudulent credit card transaction detection using soft computing techniques," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 23, no. 3, pp. 1634–1642, Sep. 2021, doi: 10.11591/ijeecs.v23.i3.pp1634-1642.
- [9] P. H. D. Abd Samad, S. Mutalib, and S. Abdul-Rahman, "Analytics of stock market prices based on machine learning algorithms," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 16, no. 2, pp. 1050–1058, 2019.
- [10] M. A. Muslim, Y. Dasril, M. Sam'an, and Y. N. Ifriza, "An improved light gradient boosting machine algorithm based on swarm algorithms for predicting loan default of peer-to-peer lending," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 28, no. 2, pp. 1002–1011, 2022.
- [11] A. K. Bitto et al., "CryptoAR: scrutinizing the trend and market of cryptocurrency using machine learning approach on time series data," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 28, no. 3, pp. 1684–1696, 2022.
- [12] S. Ghorbany, S. Yousefi, and E. Noorzai, "Evaluating and optimizing performance of public-private partnership projects using copula Bayesian network," *Eng. Constr. Archit. Manag.*, no. ahead-of-print, 2022.
- [13] Y. Wang and R. L. K. Tiong, "Public-Private Partnership Contract Failure Prediction Using Example-Dependent Cost-Sensitive Models," *J. Manag. Eng.*, vol. 38, no. 1, p. 4021079, 2022.
- [14] Y. Wang, Z. Shao, and R. L. K. Tiong, "Data-driven prediction of contract failure of public-private partnership projects," *J. Constr. Eng. Manag.*, vol. 147, no. 8, p. 4021089, 2021.
- [15] M. Eskandari, M. Taghavifard, I. R. Vanani, and S. S. G. Noori, "An Intelligent Hybrid Model for Determining Public-Private Partnership in Iranian Water and Wastewater Industry Based on Collective Tree Algorithms," *J. Water Wastewater*, vol. 32, no. 1, pp. 69–90, 2021.
- [16] Y. Wang, Z. Shao, and R. L. K. Tiong, "Data-driven prediction of contract failure of public-private partnership projects," *J. Constr. Eng. Manag.*, vol. 147, no. 8, p. 4021089, 2021.
- [17] L. L. Martin, "Making sense of public-private partnerships (P3s)," *J. Public Procure.*, 2016.
- [18] G. Y. Debela, "Critical success factors (CSFs) of public-private partnership (PPP) road projects in Ethiopia," *Int. J. Constr. Manag.*, vol. 22, no. 3, pp. 489–500, 2022.
- [19] F. Calabrò and L. Della Spina, "The public-private partnership for the enhancement of unused public buildings: an experimental model of economic feasibility project," *Sustainability*, vol. 11, no. 20, p. 5662, 2019.
- [20] D. Albalate, G. Bel, and R. R. Geddes, "Recovery risk and labor costs in public-private partnerships: Contractual choice in the US water industry," *Local Gov. Stud.*, vol. 39, no. 3, pp. 332–351, 2013.

- [21] C. B. Casady, K. Eriksson, R. E. Levitt, and W. R. Scott, "(Re) defining public-private partnerships (PPPs) in the new public governance (NPG) paradigm: an institutional maturity perspective," *Public Manag. Rev.*, vol. 22, no. 2, pp. 161–183, 2020.
- [22] C. Oliveira Cruz and J. Miranda Sarmiento, "Theory on PPP renegotiations," in *The renegotiations of public private partnerships in transportation* Springer, 2021, pp. 19–27.
- [23] M. Vaudevan, R. S. Narayanan, S. F. Nakeeb, and Abhishek, "Customer churn analysis using XGBoosted decision trees," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 25, no. 1, pp. 488–495, 2022, doi: 10.11591/ijeecs.v25.i1.pp488-495.
- [24] R. Meenal, P. A. Michael, D. Pamela, and E. Rajasekaran, "Weather prediction using random forest machine learning model," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 22, no. 2, pp. 1208–1215, 2021, doi: 10.11591/ijeecs.v22.i2.pp1208-1215.
- [25] X. Hao, G. Zhang, and S. Ma, "Deep learning," *Int. J. Semant. Comput.*, vol. 10, no. 03, pp. 417–439, 2016.