IPO Performance Prediction during Covid-19 Pandemic in Indonesian Using Decision Tree Algorithm

Arianto Muditomo¹, Ajar Susanto Broto²
¹,²School of Business, IPB University, Indonesia
*Corresponding Author: ariantomuditomo@apps.ipb.ac.id

Abstract
The purpose of this study was to explain the IPO underpricing phenomenon and to find out whether the decision tree algorithm model was able to predict the IPO performance during the Covid-19 pandemic in the Indonesian capital market. The model developed uses the IPO performance classification target variables, namely overpricing, zero, underpricing level-1 or underpricing level-2. Through the modeling of the decision tree algorithm using 149 IPO action data for 2017-2019 and tested on 45 IPO action data in 2020, the results of the study found that the decision tree algorithm was able to explain IPO performance based on the specified classification range. The use of the decision tree algorithm model to explain the IPO performance can be an alternative to the linear regression econometric model that has been widely used in previous studies to provide input for investors in making investment decisions.

Keywords : Covid-19; Decision Tree; Indonesia; IPO underpricing
JEL Classification : G10, G11

1. INTRODUCTION

Indonesia is one of the most frequent IPO performance during the Covid-19 pandemic of 2020 in South-East Asia, involving 46 companies, with the total cost of raising funding of $385 million despite being less than half the sum for the same period in the previous year (Sari, 2020). As financial-sector academics, we are interested in investigating IPO performance because it is never the case in recent decades following the global Covid-19 pandemic outbreak.

The current global stock market situation shows that the financial markets have been seriously affected by the pandemic and IPO underpricing have occurred worldwide, including in the US capital market, which is known for its high effectiveness. The excess return on the first day is known as "the secret of the IPO underpricing rate" (Han, 2016). It shows three points in particular. First, an initial short-term public offering that is underpricing, or an IPO has an abnormal return on the first day. Second, in the long term, it has a high price. Third, hot market issues emerge. Until now, many researchers and analysts of stock markets worldwide remain attracted to the IPO underpricing.
phenomenon.

Throughout the history of Indonesia’s stock market since the reactivation of a stock market by Government of the Republic of Indonesia from 1977 to 2020, there were 835 IPOs with 26 initial returns, 553 underpricing, 81 times overpricing have occurred (TICMI, 2020), 153 IPOs where data were not recorded on the ‘closing price’ (1977-1992), 11 IPO with no data of ‘closing price’ (1994-1999), 11 IPO where data on the ‘IPO price’ have not been found (1996-2005).

In the meantime, based on our analysis of 14 studies on the IPO underpricing phenomenon on the Indonesian stock exchange (IDX) over the years 2000-2019, most studies examined factors which impact underpricing (Astuti, 2020; Indriani & Marlia, 2014; Inkasari, 2019; Iryma Maygista, Indah Mustikowati, & Mardiana Firdaus, 2020; Lestari, Hidayat, & Sulasmiyati, 2015; S. T. A. Manurung & Nuzula, 2019; Munawaroh, 2019; Nathan & Scobell, 2012; Sembiring, Rahmawati, & Kusumawati, 2018; Solida, Luthan, & Sofriyeni, 2020; Sopiyani, 2019), comparison of the first day IPO closing price with the company appraisal (Herawati, Achsani, Hartoyo, & Sembel, 2016; Satriyo, 2018), and comparison of initial returns between stocks of the industrial sector (A. H. Manurung, Juwono, & Siswanti, 2019).

2. LITERATURE REVIEW

IPO Underpricing phenomenon

In this analysis, the decision tree algorithm model was developed by combining quantitative and qualitative methods for determining the IPO’s underpricing rate and identifying the key features of the underpricing phenomenon in the IPO. At the data preparation stage, qualitative methods are performed by setting up categories/space/subspace using the training data gathered to establish a classification rule model. In addition, quantitative methods are used to create a model of decision tree algorithm (Han, 2016). This research focuses on the 2017-2020 IPO actions data where the Indonesian capital market had 195 IPO actions with initial return, as shown in Table 1 below.

Table 1. Initial return (IR) IPO actions in the 2017-2020 Indonesian capital market.

<table>
<thead>
<tr>
<th>Year - Board</th>
<th>Number of companies</th>
<th>The average of IR (%)</th>
<th>Max IR (%)</th>
<th>Min IR (%)</th>
<th>Deviation Standard of IR (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tahun - Papan</td>
<td>Jumlah Perusahaan</td>
<td>Rata-rata IR (%)</td>
<td>Max IR (%)</td>
<td>Min IR (%)</td>
<td>Standard Deviasi IR (%)</td>
</tr>
<tr>
<td>2017 Development Board</td>
<td>30</td>
<td>46.1%</td>
<td>86.7%</td>
<td>-3.3%</td>
<td>28.0%</td>
</tr>
<tr>
<td>Main Board</td>
<td>7</td>
<td>29.4%</td>
<td>87.1%</td>
<td>-9.0%</td>
<td>39.0%</td>
</tr>
<tr>
<td>2018 Development Board</td>
<td>41</td>
<td>50.8%</td>
<td>70.0%</td>
<td>-35.7%</td>
<td>25.1%</td>
</tr>
<tr>
<td>Main Board</td>
<td>16</td>
<td>43.7%</td>
<td>70.0%</td>
<td>-14.3%</td>
<td>24.5%</td>
</tr>
<tr>
<td>2019 Development Board</td>
<td>44</td>
<td>46.4%</td>
<td>70.0%</td>
<td>-70.3%</td>
<td>30.6%</td>
</tr>
<tr>
<td>Main Board</td>
<td>11</td>
<td>41.6%</td>
<td>70.0%</td>
<td>-1.7%</td>
<td>26.2%</td>
</tr>
<tr>
<td>2020</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
These data show that 182 IPO underpricings occurred during the research period, where population data reflect all the activities on the Indonesian capital market. As common knowledge IPO can be divided according to several categories, such as: IPO activity, pricing, and IPO allocation (Ritter dan Welch (2002) in

As has already been pointed out, the study of previous IPO performance research in pricing and allocation categories can be divided into (W. Perera & Kulendran, 2016): the short-term underpricing phenomenon is a positive rate of return on the first day of listing, in which the closing price is greater than the price of the first bid/issue prize, the phenomenon of long-term underperformance means the acquisition of a negative rate of return at a certain period of time with the sale price lower than the initial price, the market 'hot issue' phenomenon is a short-term cyclical activity of 'underpricing,' perceived to be a continuation of the short-term phenomenon of "underpricing.". regarding its emphasis for research of short-term IPO performance, a brief description of the short-term 'underpricing' phenomenon of the IPO share is given. The IPO covers 3 directly involved parties, namely the issuer as the party that needs funds, the contractor and investor as the fund supplier. In this situation, the underwriter shall take a stance based on a 'firm commitment'. This means that the underwriter can purchase the entire issue and then sell it at any risk, or "Best efforts" meaning that the issuer will assume the loss equal to the negotiated fixed price, and that the issuer will have a "underwriting-spread" to cover incurred costs.

Various causes and explanations for the phenomenon of "underpricing" stock performance can be found in (K. L. W. Perera, 2014) where a range of theories of short-term "underpricing" (hypothesis) have been outlined as shown in Figure 1.
Among all the different theories listed above, they can be categorized as 'asymmetric' information (i.e. signalling, winner's curse, market feedback (book building), agency, investment bankers monopsony power, bandwagon and ownership dispersion), and 'Symmetrical' details, i.e. the lawsuit prevention hypothesis, the internet bubble and the trading volume. These two categories require further clarification, which focuses on 'agency disputes and actions'.

Based on the variety of theories and hypotheses, we conclude that information is an important factor determining the short-term underpricing phenomenon of IPO shares. The variables that can explain the phenomenon of 'underpricing' can be differentiated based on the root characteristics, as seen in Figure 2 below (W. Perera & Kulendran, 2016).

![Figure 2](image-url)

**Figure 2.** Variety of characteristics that have an impact on short-term IPO performance

Baba & Sevil (2020) also explains that IPO underpricing is affected by factors including company characteristics (size, profitability, profit rate) and offering characteristics (number of shares offered, share of the IPO bid, IPO price, amount of funds raised) and market sentiment (market performance before the IPO date).

Based on the analysis above, the predictor variables are measured because of (1) the accessible information for potential investors and (2) the coverage of company characteristics, offering characteristics and market characteristics.
Decision Tree Algorithm Model for IPO Research

The decision tree algorithm is a traditional data mining algorithm which investigates findings based on examples. The focus of this algorithm is to consider the classification rules in a decision tree of many examples without order and without rules. These classification rules are frequently used to establish predictive models or classification that can classify unknown patterns and identify it.

Researchers also find it confusing on when using linear regression algorithms and when using decision tree algorithms to make predictions. Joshi (2017) notes that linear regression is not suitable for classification because it relies on linearly interconnected data, while decision tree is a compatible algorithm for classification. Most IPO performance research has been performed on linear regression models, but in this study we find previous research using the decision tree algorithm (Basti, Kuzey, & Delen, 2015; Chen, Chen, & Cheng, 2010; Chen & Cheng, 2012; Han, 2016; Luque, Quintana, & Isasi, 2012; Quintana, Luque, Valls, & Isasi, 2012; Quintana, Sáez, & Isasi, 2017). We find that machine learning algorithms (the decision tree is known in this case as supervised learning) contain fewer error predictions than linear regression.

The decision tree model offers many benefits, including rules generated by the decision tree are easily understood, higher efficiency, suitable for samples with large amounts of data, and higher accuracy in classification (Han, 2016).

Because of these considerations, we agreed to use a two-stage developmental model for a decision tree: firstly to use the IPO action data set for training (we used the IPO 2017/2019) to plan and define the decision tree; and secondly, to use the test data set (we use the IPO 2020 action data) for classification using the established model. The different models tested were decision tree (simple), random forest and boosted gradient tree with the best test indicators based on the minimum model error. The different models tested were decision tree (simple), forest random and gradient boosted tree, with the best indicators based on the model's minimum error.

3. METHODS, DATA AND ANALYSIS

Method

This research examines a model of a decision tree algorithm to decide whether an IPO performance would lead to 'underpricing,' 'zero' or 'overpricing.' IPO performance is measured by the IPO_Perf formula = (first day price-initial offer price)/initial offer price. Moreover, the predictor variable is determined based on three characteristics (Baba & Sevil, 2020; W. Perera & Kulendran, 2016) including (a) for the characteristics of the company, recording board proxy is chosen to represent the criteria for the operational life of the company (recording operating income), operating profit, issuance of audited financial statements and financial measuring activities of the company (IDX, 2019); (b) for offering characteristics, the selected variables are proxy of raised funds (Fund_R), share of shares sold (Pct_IPO), and total shares offered (Shares_Off), which represents the demand side and the offer price (Offer_P) representing the supply side; (c) for market characteristics, a proxy for IPO date (List_date) is chosen to reflect the seasonal model of the market (hot-cold issue). These variables are chosen to accord with (Chen & Cheng, 2012; Han, 2016).

Data

Data used in this analysis are population data with a set of indicators for each
variable based on publications (BEI, 2020; TICMI, 2020). The data is then separated into training data and data sets linearly. The training data is the IPO performance data before the pandemic period 2017-2019 and the evaluation data are IPO performance data after the 2020 pandemic.

We pre-processed the data by identifying each variable in the processed data set before modelling and analysis.

**Table 2. Variables for the analysis.**

<table>
<thead>
<tr>
<th>No</th>
<th>Variables</th>
<th>Definition of variables</th>
<th>Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>IPO_Perf (target)</td>
<td>IPO performance = (1st day opening price-initial offer price)/initial offer price</td>
<td>0 - (zero IR); 1 - (IR up to 44% underpricing; 2 - (IR &gt; 44% underpricing; -1 - (overpricing IR) - 44% is the mean IR from data set training</td>
</tr>
<tr>
<td>2</td>
<td>Board</td>
<td>Boards for listings on IPO performance</td>
<td>1 - (Main Board); 2 - (Development Board); 3 - (Acceleration Board)</td>
</tr>
<tr>
<td>3</td>
<td>Fund_R</td>
<td>Funds raised</td>
<td>Continuous (min-max)</td>
</tr>
<tr>
<td>4</td>
<td>Pct_IPO</td>
<td>Shares sold = offered shares / issued shares</td>
<td>Continuous (min-max)</td>
</tr>
<tr>
<td>5</td>
<td>Shares_Off</td>
<td>The number of shares offered</td>
<td>Continuous (min-max)</td>
</tr>
<tr>
<td>6</td>
<td>Offer_P</td>
<td>IPO offering price</td>
<td>Continuous (min-max)</td>
</tr>
<tr>
<td>7</td>
<td>List_date</td>
<td>IPO date - converted to a quarter classification</td>
<td>1 - (Q1); 2 - (Q2); 3 - (Q3); 4 - (Q4)</td>
</tr>
</tbody>
</table>

**Analysis**

For data analysis, the automotive model feature in the Rapidminer ver application, 9.8 001 is used to simulate a model with training data and to measure error to get the best model with minimum errors, with performance is further checked by test data.

The auto model function is used to find the best decision tree algorithm model from the 3 models available: decision tree (simple), random forest and gradient boosted tree. In addition, the best model is used to create and use a predictive model for testing data and to interpret the prediction results.

**4. RESULTS**

This study is to explain the IPO underpricing theory in the Indonesian capital market during the Covid-19 pandemic, following the short-term IPO underpricing theory framework that addresses the characteristics of issues, companies and markets (Perera 2016). In addition, a model of prediction with decision tree algorithms is developed to deal with nonlinearity problems (Baba & Sevil, 2020; Han, 2016).

**Descriptive Analysis**

The descriptive statistics of the training data set are listed in Table 3 below.
Table 3. Descriptive statistics

<table>
<thead>
<tr>
<th>No</th>
<th>Variables</th>
<th>Data type</th>
<th>Min (Least*)</th>
<th>Max (Most*)</th>
<th>Average (Values*)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>IPO_Perf (target)</td>
<td>Real</td>
<td>22,174,000</td>
<td>4,000,000,000</td>
<td>734,286,585</td>
</tr>
<tr>
<td>2.</td>
<td>Board</td>
<td>Real</td>
<td>0.005</td>
<td>0.879</td>
<td>0.247</td>
</tr>
<tr>
<td>3.</td>
<td>Fund_R</td>
<td>Integer</td>
<td>100</td>
<td>12,100</td>
<td>527.17</td>
</tr>
<tr>
<td>4.</td>
<td>Pct_IPO</td>
<td>Integer</td>
<td>1</td>
<td>4</td>
<td>2.89</td>
</tr>
<tr>
<td>5.</td>
<td>Shares_Off</td>
<td>Integer</td>
<td>-1</td>
<td>2</td>
<td>1.59</td>
</tr>
<tr>
<td>6.</td>
<td>Offer_P</td>
<td>Real</td>
<td>16,200,000,000</td>
<td>4,764,375,000,000</td>
<td>275,223,575,557</td>
</tr>
<tr>
<td>7.</td>
<td>List_date*</td>
<td>Polynomial</td>
<td>Main (34)</td>
<td>Board Development Board (115)</td>
<td>Development Board (115), Main Board (34)</td>
</tr>
</tbody>
</table>

Table 3 illustrates that all training data sets can be assigned a classification value according to their definition.

**Model Developments**

The classification capacity of the constructed model is shown as a minimum error. We divided it into 149 training data (IPO 2017-2019) and 46 test data from the 185 data (IPO 2020 actions). The data collection for statistical training includes 115 IPO actions 2017-2019 on the “Development Board” and 34 IPO actions on the “Main Board”. The correlation weights of each input variable to the target variable are as follows:

![Figure 3. Correlation Weights of Input Variables to Target Variables (generated from Rapid Miner data processing).](image)

These results show that Fund R has the highest correlation weight to the target variable IPO_Perf class followed by Offer_P, Shares_Off, Pct_IPO, Board and List_Date. This provides an initial prediction of the established decision tree model that fits the
above correlation weights.

In addition, Model performance testing (decision tree (simple), random forest, and gradient-boosted tree) produces minimal error (root mean squared error, absolute error, relative errors and squared error) for the Decision Tree (simple) model, as explained in the following table.

<table>
<thead>
<tr>
<th>Model</th>
<th>RMSE</th>
<th>Absolute</th>
<th>Relative</th>
<th>Squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decision Tree</td>
<td>0.62</td>
<td>0.46</td>
<td>0.24</td>
<td>0.49</td>
</tr>
<tr>
<td>Random Forest</td>
<td>0.68</td>
<td>0.55</td>
<td>0.33</td>
<td>0.50</td>
</tr>
<tr>
<td>Gradient Boosted Trees</td>
<td>0.70</td>
<td>0.56</td>
<td>0.32</td>
<td>0.54</td>
</tr>
<tr>
<td>Generalized Linear Model</td>
<td>0.63</td>
<td>0.51</td>
<td>0.29</td>
<td>0.52</td>
</tr>
</tbody>
</table>

From the illustration of the table above, the Decision Tree (simple) algorithm produces minimal errors on all indicators of measured error, in particular when compared to a generalized linear model (representing linear regression). We therefore decided to use a (simple) algorithm model for the decision tree model.

Given that there was no IPO activity on the Acceleration Board in the Training Dataset Period for 2017-2019, the decision tree model did not have the target classification of the Acceleration Boards. Therefore the final model for the decision tree is shown in Figure 4.

Figure 4. IPO Performance Prediction Model (generated from Rapid Miner data processing)
Based on the final decision tree model, the root node of the model is the Fund-Raised Index (Fund_R), meaning that the IPO performance is primarily determined by the amount of funds raised by the issuer. If IPO action hits IDR 800 billion in fundraising, investors have a significant chance to get an initial return of over 44%. Whereas, if the IPO collected funds are greater than Rp. 800 billion, the listing board consideration is critical. This means that if the IPO is listed on the Development Board, investors would have a significant opportunity to gain an initial return of 44%. If the IPO is on the 'main board' then investors are required to take into account the percentage of shares sold to the public against all listed shares.

If publicly sold shares are less than 16 percent, investors should consider buying the issuer's shares in the secondary market after the IPO, since the initial return on shares is expected to be below the offering price ("overpriced").

Meanwhile, if shares sold to the public are > 16 percent, the initial return will be determined by the number of shares offered to the public. This would lead to an initial return of up to 44% if the amount of shares sold is over 24 billion shares.

When the model is tested on the 2020 IPO action test data, the error matrix is structured as follows:

<table>
<thead>
<tr>
<th>Prediction</th>
<th>-1</th>
<th>0</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actual</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-1</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>24</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td>15</td>
</tr>
</tbody>
</table>

Note: 5 IPO actions on the 'acceleration board' in 2020 are unpredictable.

5. CONCLUSIONS, LIMITATIONS, AND RECOMMENDATIONS

Conclusions
Based on the results of the aforementioned analysis with the algorithm model decision tree, we conclude that the prediction of underpricing of IPO performance can explain the IPO underpricing phenomenon and complement the linear regression model commonly used in previous studies.

In predicting IPO performance of underpricing, the decision-tree algorithm model can deal with non-linearity issues, address nominal variables and contribute to the existing knowledge by overcoming limitation of linear prediction models which cannot accommodate nominal variables. Therefore, the established model can be an alternative investment decision-making model for investors based on general knowledge that can be accessed from a prospectus or the stock exchange authority.

By considering many IPO metrics, such as the number of shares sold to the public, the percentage of the number of shares sold to all issued shares, the IPO offering price, raised funds and the listing board, investors can predict the initial returns that they will receive while participating in the company's IPO action.

Machine study algorithms and rapid-miner applications are the key characteristics
of this study as they encourage the interest of junior researchers who are more interested in machine learning than conventional statistical learning. However, the shift in the interest of these junior researchers hopefully will not diminish their enthusiasm in econometrics in general and financial econometrics in particular.

**Limitations and Recommendations**

Although this study shows that the simple decision tree algorithm can achieve better predictive performance, it is suggested in future studies to extend the time for the training model data set so that the model can better capture population behaviour. Input variables can be added with an in-depth analysis of variable characteristics influencing short-term IPO performance (supply characteristics, company characteristics and market characteristics), as illustrated in Figure 2.

**REFERENCES**


