

Development of a Classification Model for Underpriced Issuers Using Machine Learning Algorithms

Defitra Hidayatullah ^{1*}, Ihsan Jatnika ^{1*}

* Manajemen Sistem Informasi, Universitas Gunadarma
defitra.hidayatullah@yahoo.com ¹, ihsan@staff.gunadarma.ac.id ²

Article Info

Article history:

Received 2025-06-06

Revised 2025-07-07

Accepted 2025-07-19

Keyword:

Initial Public Offering,
Underpricing,
CRISP-DM,
Machine Learning,
Confusion Matrix.

ABSTRACT

This study develops a classification model to identify underpriced Initial Public Offering (IPO) issuers so that it can help investor decision-making. Using the CRISP-DM methodology, this research uses a sample of 209 non-banking IPO issuers of the OJK E-IPO platform (since establishment until December 31, 2024). In addressing the problem of class imbalance (161 underpriced and 48 not underpriced), SMOTE was used. The model utilizes nine features: Year-on-Year, IHSG, IPO price, ratio of shares issued, age of the firm, size of the firm, sales growth, Return on Assets (ROA), Debt to Equity Ratio (DER), and Asset Turnover Ratio (ATO). Seven classifier algorithms were compared based on accuracy, precision, sensitivity, recall, F1-score, and AUC. Random Forest had the best performance with 89.2% accuracy, 88.9% Macro Average F1-score, and an AUC of 0.946. The findings suggest that the Random Forest model accurately identifies underpriced IPO issuers as a good investment decision-making tool. This research demonstrates that machine learning concepts can be implemented to classify underpriced issuers in Indonesia, continuing previous studies that contributed to understanding the correlation and significance of certain variables to underpricing.



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

I. PENDAHULUAN

Statistik pasar modal menunjukkan jumlah investor pasar modal semakin meningkat dari tahun ke tahun. Hal ini menunjukkan antusiasme masyarakat dalam berinvestasi, khususnya di pasar modal Indonesia. Investasi dapat diartikan sebagai aktivitas menunda konsumsi saat ini untuk dialokasikan pada aset produktif dalam jangka waktu tertentu [1]. Sedangkan pasar modal dapat diartikan sebagai tempat berbagai instrumen jangka panjang yang dapat ditransaksikan [2]. Salah satu pilihan investasi di pasar modal adalah pembelian saham IPO. IPO atau *Initial Public Offering* merupakan serangkaian proses yang harus dilalui perusahaan yang ingin menjual sahamnya kepada masyarakat sebelum saham tersebut dapat ditransaksikan di bursa saham. Menurut Sugiyanto *et al*, keputusan untuk membeli saham IPO memiliki risiko yang tergolong tinggi [3]. Salah satu informasi yang cukup penting dalam membeli saham IPO adalah informasi *underpricing* emiten. *Underpricing* merupakan selisih harga antara harga IPO dan harga *listing*.

Harga IPO ditentukan oleh kesepakatan emiten dan *underwriter*. Sedangkan harga *listing* terbentuk dari mekanisme pasar [4].

Underpricing merupakan fenomena yang umum terjadi pada saat perusahaan menawarkan sahamnya melalui IPO. Suatu saham dikatakan *underprice* apabila memiliki harga IPO yang lebih rendah daripada harga sebenarnya dan dijual lebih tinggi pada saat hari pertama *listing* di bursa [5]. Emiten dengan harga IPO yang lebih rendah daripada harga aslinya cenderung menarik bagi investor. Hal ini dikarenakan investor berpotensi mendapatkan *initial return* berupa selisih harga beli saham IPO dan harga jual di bursa [6]. *Underpricing* tidak hanya penting bagi investor, melainkan juga bagi *underwriter*. Minat yang tinggi terhadap suatu saham IPO dapat menjadi tanda positif untuk memastikan saham yang ditawarkan terjual habis dan *underwriter* tidak menanggung saham yang tidak terjual [7]. Menurut Maylani *et al*, terdapat dua faktor yang secara umum berpengaruh terhadap tingkat *underpricing* suatu emiten pada saat IPO, yaitu faktor finansial dan faktor non-finansial [8].

Penelitian mengenai *underpricing* sudah banyak dilakukan oleh para peneliti terdahulu. Misalnya penelitian yang dilakukan oleh Apriliani dan Bustamam (2022) yang meneliti mengenai pengaruh *financial leverage*, ukuran perusahaan, dan kepemilikan institusional terhadap *underpricing*. Hasilnya menunjukkan bahwa *financial leverage*, ukuran perusahaan, dan kepemilikan institusional secara simultan maupun parsial tidak berpengaruh terhadap *underpricing* saham syariah saat *Initial Public Offering* (IPO) di BEI pada masa *Covid-19* [9]. Sejalan dengan hasil tersebut, Rohman dan Fitriati (2024) mengungkapkan bahwa umur perusahaan, ukuran perusahaan, ROA, dan DER secara simultan tidak berpengaruh terhadap tingkat *underpricing* saham pada perusahaan jasa yang melakukan penawaran umum perdana (IPO) di Bursa Efek Indonesia [10]. Hasil penelitian Hakim dan Fauzan (2023) menyatakan bahwa *return on asset* (ROA), *return on equity* (ROE), *debt to equity ratio* (DER), ukuran perusahaan, rencana penggunaan dana, *underwriter*, umur perusahaan, dan persentase saham ditawarkan mempengaruhi *underpricing* secara simultan dan hanya ukuran perusahaan yang berpengaruh signifikan positif terhadap *underpricing* secara parsial (studi kasus pada emiten yang IPO periode 2020-2022) [11]. Sedangkan penelitian yang dilakukan oleh Andriani *et al* (2022) yang mengungkapkan bahwa *current ratio* berpengaruh negatif signifikan, sedangkan *return on asset* dan nilai tukar tidak berpengaruh negatif signifikan terhadap *underpricing* pada perusahaan manufaktur yang IPO tahun 2015-2020. Ketiga variabel tersebut secara simultan berpengaruh signifikan terhadap *underpricing* pada perusahaan manufaktur yang IPO tahun 2015-2020 [12]. Suryawan dan Yuniarta (2023) mengonfirmasi bahwa *financial distress*, *financial leverage*, umur perusahaan, dan tingkat inflasi secara bersama-sama tidak mempengaruhi *underpricing* saham IPO pada 2019-2021 [13]. Adapun penelitian oleh Sonia *et al* menyimpulkan bahwa profitabilitas dan umur perusahaan tidak berpengaruh terhadap *underpricing*. Sedangkan variabel *leverage* berpengaruh positif terhadap *underpricing* (studi kasus pada perusahaan yang melakukan IPO di BEI Tahun 2020-2023)[14]. Hasil ini sedikit berbeda dengan penelitian yang dilakukan oleh Putri dan Ekaningtias (2024) dimana hasil penelitian menunjukkan bahwa variabel profitabilitas berpengaruh pada tingkat *underpricing* saham *initial public offering*, sedangkan *firm size* perusahaan dan *financial leverage* tidak memiliki pengaruh pada tingkat *underpricing* saham [15].

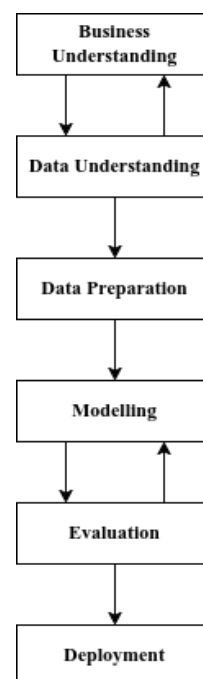
Perbedaan mendasar antara penelitian ini dan penelitian sebelumnya adalah penerapan algoritma *machine learning* untuk mengidentifikasi pola emiten yang *underpricing*. *Machine learning* merupakan cabang ilmu kecerdasan buatan yang memiliki kemampuan untuk mempelajari data dan mengidentifikasi pola-pola tertentu [16]. Penelitian ini menggunakan salah satu jenis *machine learning*, yaitu *supervised learning* yang menggunakan data historis berlabel dalam membuat model dan memprediksi data baru yang tidak diberi label, salah satu contoh algoritma yang termasuk *supervised learning* adalah *K-Nearest Neighbor* [17]. Algoritma *supervised learning* dapat diimplementasikan

untuk membuat model klasifikasi. Klasifikasi adalah teknik dalam data mining yang mengelompokkan data ke dalam satu kelas tertentu berdasarkan fitur yang dimiliki [17].

Algoritma yang digunakan pada penelitian ini diantaranya *Naive Bayes*, *K-Nearest Neighbors*, *SVM*, *Decision Trees*, *Logistic Regression*, *Neural Networks* dan *Random Forest*. *K-Nearest Neighbors* merupakan algoritma yang didasarkan pada prinsip bahwa titik yang serupa cenderung memberikan hasil yang serupa pula. Cara kerja KNN adalah dengan menghitung jarak antara titik data baru dan tetangga (*neighbor*) terdekatnya dalam ruang fitur [18]. *Neural Network* merupakan model yang didasarkan pada cara kerja otak manusia yang terdiri dari banyak neuron [19]. *Decision Tree* merupakan algoritma yang menghasilkan pohon keputusan dalam mengklasifikasikan data. Algoritma *random forest* merupakan pengembangan dari algoritma *decision tree*. *Random Forest* (RF) adalah algoritma yang mampu meningkatkan tingkat akurasi dalam membangkitkan atribut untuk setiap *node* yang dilakukan secara acak [20]. Adapun *SVM* merupakan algoritma yang berusaha mencari *hyperplane* paling optimal sebagai pemisah dua kelas pada ruang input [21].

II. METODE

Sebagaimana alur kerja pada gambar 1, metode penelitian didasarkan pada kerangka kerja CRISP-DM. CRISP-DM telah menjadi standar industri dalam proyek *Data Mining*. Penggunaan CRISP-DM memastikan penelitian dapat berjalan dengan terstruktur dan terorganisir.



Gambar 1. Kerangka Kerja CRISP-DM

Data yang digunakan pada penelitian ini adalah data sekunder yang diolah dari laporan keuangan emiten dan

informasi penawaran umum yang terdapat di dalam prospektus.

Terdapat 9 (sembilan) fitur yang digunakan, yaitu *IHSG year on year* pada saat penawaran umum, harga IPO, persentase saham yang ditawarkan, umur emiten, ukuran emiten, pertumbuhan penjualan/pendapatan, *return on asset*, *debt to equity ratio*, dan *asset turnover*. Semua fitur ini diperoleh melalui prospektus emiten, kecuali *IHSG year on year* pada saat penawaran umum yang diperoleh dari *Yahoo Finance*. Pemberian label dilakukan berdasarkan tingkat *underpricing* emiten. Harga *listing* emiten yang digunakan untuk menghitung tingkat *underpricing* adalah harga penutupan hari pertama emiten *listing* di Bursa yang diperoleh melalui aplikasi stockbit.

Adapun kriteria data yang digunakan adalah sebagai berikut:

- 1) Emiten bukan merupakan lembaga perbankan.
- 2) Emiten melakukan penawaran umum dan terdaftar di platform E-IPO OJK sejak pertama kali website tersebut diluncurkan hingga 31 Desember 2024.

Berdasarkan kriteria tersebut, data emiten yang digunakan dalam penelitian ini adalah sebanyak 209 data.

A. Business Understanding

Selama ini, investor mengandalkan analisis rasio keuangan sebagai salah satu aspek yang diperhatikan dalam mengklasifikasikan emiten yang *underpricing*. Seiring dengan mudahnya pasar modal diakses oleh berbagai pihak dengan beragam latar belakang, analisis ini cukup sulit dilakukan dan bukan tidak mungkin menghasilkan kesimpulan yang bias terutama bagi investor pemula. Hal ini dikarenakan interpretasi terhadap rasio keuangan bersifat relatif. Jika diperhatikan lebih dalam, analisis ini pada dasarnya membentuk pola tertentu yang menjadi karakteristik emiten yang *underpricing*. Oleh karena itu, algoritma *machine learning* dapat diterapkan untuk mengidentifikasi pola tersebut dan diharapkan memudahkan investor dalam pengambilan keputusan investasi, dalam konteks pembelian saham IPO.

B. Data Understanding

Berdasarkan kriteria data yang telah ditetapkan, diperoleh data sebanyak 209 emiten telah melakukan IPO dimana sebanyak 161 emiten *underpricing* dan 48 emiten tidak *underpricing*. Informasi mengenai emiten yang IPO disajikan secara lengkap dalam bentuk prospektus. Prospektus dapat diunduh oleh investor melalui situs E-IPO. Di dalam prospektus terkandung informasi penawaran umum, ringkasan emiten, pihak-pihak terkait hingga laporan keuangan emiten. Untuk membangun model, diperlukan fitur dan label. Pemberian label didasarkan pada tingkat *underpricing* emiten. Tingkat *underpricing* emiten merupakan rasio yang menggambarkan selisih harga emiten pada saat IPO dan *listing* hari pertama. Rumus untuk menghitung tingkat *underpricing* adalah sebagai berikut:

$$\text{Tingkat Underpricing} = \frac{\text{Harga Listing hari pertama} - \text{Harga IPO}}{\text{Harga IPO}} \quad (1)$$

Emiten diberi label “UNDERPRICING” jika tingkat *underpricing* > 0 . Sebaliknya, emiten diberi label “NOT_UNDERPRICING” jika tingkat *underpricing* ≤ 0 .

Fitur-fitur yang digunakan pada penelitian ini didasarkan pada relevansinya dalam merepresentasikan berbagai aspek dari emiten berdasarkan studi literatur.

- 1) Aspek emiten direpresentasikan oleh umur emiten dan ukuran emiten.
- 2) Aspek IPO direpresentasikan oleh harga IPO dan persentase saham yang ditawarkan.
- 3) Aspek pasar direpresentasikan oleh perubahan harga *IHSG year-on-year*.
- 4) Aspek profitabilitas direpresentasikan oleh *return on assets* (ROA).
- 5) Aspek solvabilitas direpresentasikan oleh *debt to equity ratio* (DER).
- 6) Aspek aktivitas direpresentasikan oleh *assets turnover* (ATO).
- 7) Aspek penjualan direpresentasikan oleh *sales growth*.

Berikut adalah tabel gambaran fitur dan label pada dataset yang digunakan dalam penelitian:

TABEL I
FITUR DAN LABEL YANG DIGUNAKAN

Fitur dan Label	Nama Variabel
IHSG year on year	ihsg_yoy
Harga IPO	ipo_price
Persentase saham yang ditawarkan	ps
Umur emiten	age
Ukuran emiten	firm_size
Pertumbuhan penjualan	sales_growth
Return on asset	roa
Debt to equity ratio	der
Asset Turnover	ato
Label underpricing	Underpricing / Not_underpricing

Berdasarkan Tabel I, fitur-fitur yang digunakan diantaranya adalah *IHSG year on year* pada saat penawaran umum, harga IPO, persentase saham yang ditawarkan, umur emiten, ukuran emiten, pertumbuhan penjualan/pendapatan, *return on asset*, *debt to equity ratio*, dan *asset turnover*. Sedangkan label yang digunakan adalah label *underpricing*.

C. Data Preparation

Tahapan persiapan data dimulai dengan melakukan input data dari prospektus ke dalam tabel. Semua fitur yang diperlukan pada umumnya telah tersedia pada ikhtisar rasio keuangan prospektus. Namun apabila tidak tersedia, perhitungan secara manual masih dapat dilakukan. Meski begitu, masih ada beberapa emiten yang tidak menyertakan salah satu fitur dan tidak pula menyertakan informasi keuangan untuk menghitung fitur tersebut. Secara total, terdapat 7 *missing value* pada data *sales_growth*. Selain itu, terdapat emiten dengan fitur yang tergolong *outlier*. Oleh

karena itu, perlu dilakukan pembersihan untuk mendapatkan data yang berkualitas.

Identifikasi *outlier* dilakukan menggunakan *boxplot* dengan pendekatan *Interquartile Range* (IQR). Batas pendeteksian *outlier* ditentukan berdasarkan rumus berikut:

$$Q1 = \text{Kuartil Pertama} \quad (2)$$

$$Q3 = \text{Kuartil Ketiga} \quad (3)$$

$$IQR = Q3 - Q1 \quad (4)$$

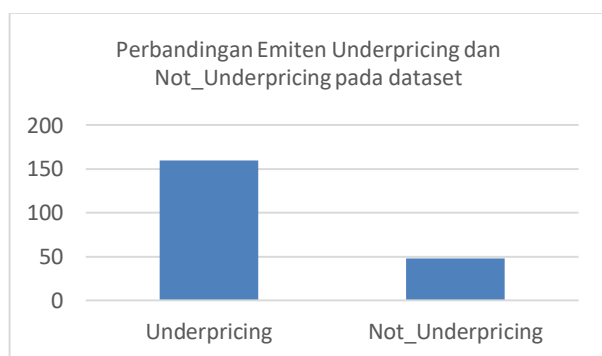
$$\text{Lower Boundary} = Q1 - (1.5 \times IQR) \quad (5)$$

$$\text{Upper Boundary} = Q3 + (1.5 \times IQR) \quad (6)$$

Setiap nilai yang berada di atas *Upper Boundary* dan di bawah *Lower Boundary* diidentifikasi sebagai *outlier*. Penanganan terhadap data yang tergolong *outlier* dilakukan dengan imputasi nilai rata-rata sesuai kelas emiten dimana nilai asal *outlier* itu sendiri tidak diikutkan dalam perhitungan rata-rata.

Langkah berikutnya adalah mengisi *missing value* dengan nilai rata-rata. Untuk menjaga konsistensi, imputasi dilakukan secara terpisah berdasarkan label kelas masing-masing. *Missing value* pada fitur emiten *underpricing* diisi dengan nilai rata-rata fitur yang sama pada kelompok *underpricing*. Sebaliknya, *missing value* pada fitur emiten yang tidak *underpricing* diisi dengan nilai rata-rata fitur yang sama pada kelompok *not_underpricing*.

Setelah data dibersihkan, berikut adalah perbandingan antara emiten yang *underpricing* dan tidak *underpricing* setelah dilakukan pembersihan data:

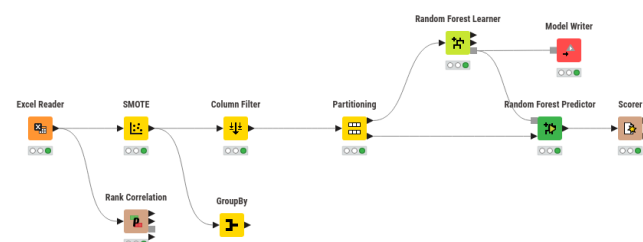


Gambar 2. Perbandingan jumlah data berdasarkan label

Diagram pada gambar 2 menunjukkan bahwa kelas *underpricing* menjadi kelas mayoritas, yaitu sebanyak 161 data. Sedangkan kelas *not_underpricing* menjadi kelas minoritas, yaitu sebanyak 48 data. Data yang *imbalance* ini akan tangani dengan SMOTE pada tahap *modelling*. Secara keseluruhan, diperoleh dataset setelah *cleaning data* sebanyak 209 data. Dataset ini siap untuk diproses pada tahap berikutnya, yaitu tahap *modelling*.

D. Modelling

Modelling dilakukan dengan bantuan aplikasi KNIME. Aplikasi ini sudah cukup mengakomodasi keperluan untuk membangun model. Berikut adalah skema *modelling* yang digunakan pada aplikasi KNIME:



Gambar 3. Skema *modelling* pada aplikasi KNIME

Sebagaimana pada gambar 3, tahap pertama adalah melakukan impor dataset dengan widget "excel reader". Tahapan kedua adalah mengukur tingkat korelasi fitur terhadap label. Tahap ketiga adalah melakukan *oversampling* terhadap data label minoritas agar data seimbang secara kuantitas menggunakan metode SMOTE. Dari proses penyeimbangan data, diperoleh sebanyak 322. Tahap keempat adalah melakukan filter untuk mengeluarkan kolom tingkat *underpricing* dari dataset. Hal ini dikarenakan kolom tingkat *underpricing* hanya digunakan untuk memberi label dan tidak relevan untuk digunakan membangun model. Tahap kelima adalah melakukan normalisasi pada data yang akan digunakan oleh algoritma yang sensitif terhadap skala fitur, yaitu Support Vector Machine (SVM), K-Nearest Neighbors (KNN), dan Neural Network. Metode normalisasi yang digunakan adalah Min-Max Scaling, yang mengubah skala fitur ke dalam rentang antara 0 dan 1. Pada tahap keenam, dataset dibagi menjadi *data training* dan *data testing* secara acak dengan skema pembagian 80:20. Dari proses *split data*, diperoleh sebanyak 257 *data training* dan 65 *data testing*. Tahap ketujuh adalah membangun model dengan serangkaian uji coba menggunakan beberapa algoritma klasifikasi, yaitu *naive bayes*, *logistic regression*, SVM, KNN, *neural network*, *decision tree* dan *random forest*. Algoritma terbaik dipilih berdasarkan beberapa metrik pada tahap berikutnya, yaitu tahap *evaluation*.

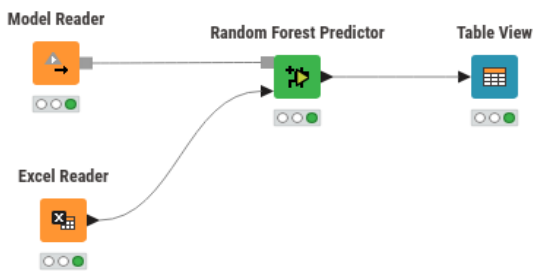
E. Evaluation

Pengujian dilakukan dengan metode *confusion matrix* terhadap data yang belum pernah diidentifikasi oleh model, yaitu *data testing*. Selain itu, metrik *accuracy*, *F1-Score*, *Recall*, *Precision*, *Sensitivity* hingga *AUC* digunakan sebagai tolok untuk perbandingan kinerja antar model.

F. Deployment

Tahap *Deployment* tidak selalu berkaitan dengan pembangunan aplikasi, melainkan juga penerapan model yang sudah dibangun untuk mengklasifikasi data pada dunia nyata. Pada penelitian ini, *deployment* yang dimaksud adalah penerapan model *random forest* untuk mengklasifikasi 8 emiten yang *listing* pada Januari 2025. Data ini merupakan

data baru dan tidak termasuk ke dalam dataset. Berikut adalah skema *deployment* yang digunakan:



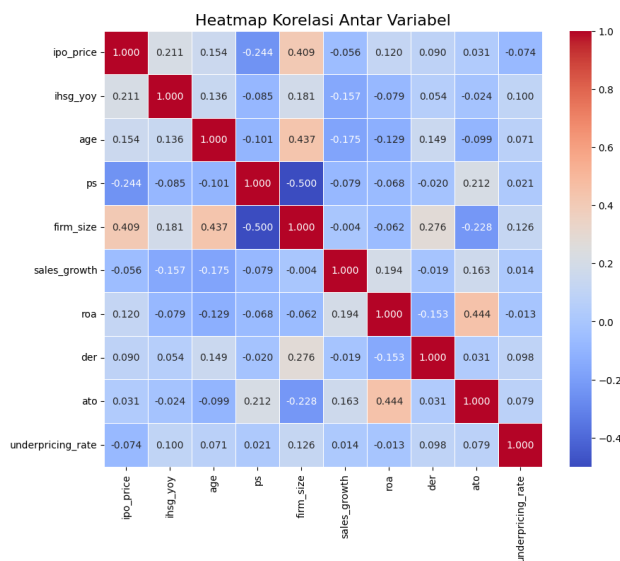
Gambar 4. Skema deployment pada aplikasi KNIME

Sebagaimana diilustrasikan pada gambar 4, tahapan *deployment* dimulai dari mengimpor model terbaik yang dihasilkan dan data emiten yang ingin diklasifikasikan. Kemudian model akan mengklasifikasi emiten dan hasil klasifikasi ditampilkan dalam tabel.

III. HASIL DAN PEMBAHASAN

A. Correlation Matrix

Analisis korelasi dilakukan untuk mengetahui kekuatan dan hubungan antar variabel. Berikut adalah *correlation matrix* yang divisualisasikan dalam bentuk *heatmap*.



Gambar 5. Correlation matrix

Pada gambar 5, tingkat korelasi fitur terhadap label tergolong lemah. Tingkat korelasi fitur terhadap tingkat *underpricing* hanya berada dalam jangkauan -0,074 hingga 0,126. Hasil ini sejalan dengan beberapa penelitian terdahulu yang menyatakan bahwa beberapa rasio keuangan memiliki tingkat korelasi yang lemah terhadap tingkat *underpricing*. Fitur ukuran emiten menjadi fitur dengan tingkat korelasi paling tinggi. Meski tergolong lemah, bukan berarti fitur-fitur ini tidak memiliki hubungan terhadap *underpricing*. Hubungan fitur-fitur terhadap tingkat *underpricing* dapat berupa hubungan yang lebih kompleks dan bukan hanya

sekadar hubungan linier biasa. Penerapan konsep *machine learning* dapat menjadi solusi untuk menemukan hubungan tersebut.

B. Hasil Evaluasi Algoritma

Analisis kinerja model diukur berdasarkan metrik Recall, Precision, Sensitivity, Specificity, F1-Score, Accuracy dan AUC. Berikut adalah tabel perbandingan evaluasi masing-masing model yang diterapkan:

TABEL II
HASIL EVALUASI RECALL, PRECISION, SENSITIVITY, DAN SPECIFICITY

Algorithm	Recall	Precision	Sensitivity	Specificity
Random Forest	0,884	0,898	0,884	0,946
Neural Network	0,803	0,819	0,803	0,892
KNN	0,767	0,793	0,767	0,892
Decision Tree	0,758	0,768	0,758	0,838
Naive Bayes	0,664	0,670	0,664	0,757
Logistic Regression	0,619	0,622	0,619	0,703
SVM	0,622	0,750	0,622	0,243

TABEL III
HASIL EVALUASI F1-SCORE, ACCURACY, COHEN'S KAPPA, DAN AUC

Algorithm	F-1 Score	Accuracy	Cohen's kappa	AUC
Random Forest	0,889	0,892	0,778	0,946
Neural Network	0,808	0,815	0,617	0,843
KNN	0,773	0,785	0,549	0,823
Decision Tree	0,761	0,769	0,523	0,731
Naive Bayes	0,666	0,677	0,333	0,673
Logistic Regression	0,620	0,631	0,241	0,632
SVM	0,529	0,569	0,217	0,519

Berdasarkan tabel II dan tabel III, dapat disimpulkan bahwa algoritma *random forest* menjadi algoritma terbaik dalam mengklasifikasi emiten *underpricing*. Algoritma ini memiliki metrik dengan nilai tertinggi yang mengindikasikan kinerjanya lebih baik dan lebih konsisten dibanding algoritma lainnya. Berikut adalah *confusion matrix* algoritma *random forest*.

TABEL IV
CONFUSION MATRIX MODEL TERBAIK (RANDOM FOREST)

is_underpricing? \ Prediction	NOT_UNDERPRICING	UNDERPRICING
NOT_UNDERPRICING	35	2
UNDERPRICING	5	23

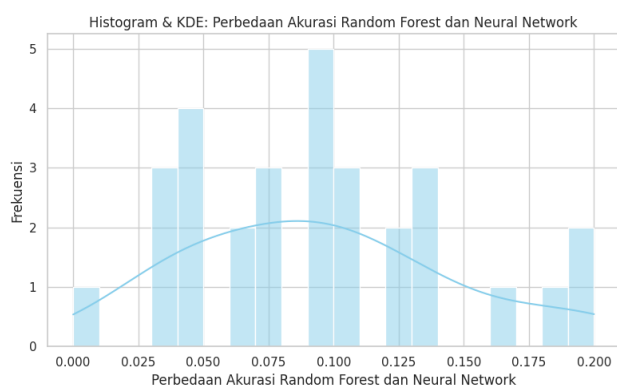
Berdasarkan hasil *confusion matrix* pada Tabel IV, diperoleh hasil evaluasi sebagai berikut:

- 1) *True Negatives* (TN): 35 emiten yang seharusnya NOT_UNDERPRICING berhasil diprediksi dengan benar sebagai NOT_UNDERPRICING.

- 2) *False Positives* (FP): 2 emiten yang seharusnya NOT_UNDERPRICING salah diprediksi sebagai UNDERPRICING. Ini mewakili kesalahan Tipe I, di mana model salah mengidentifikasi non-underpricing sebagai *underpricing*.
- 3) *False Negatives* (FN): 5 emiten yang seharusnya UNDERPRICING salah diprediksi sebagai NOT_UNDERPRICING. Ini mewakili kesalahan Tipe II, di mana model gagal mengidentifikasi emiten yang sebenarnya *underpricing*.
- 4) *True Positives* (TP): 23 emiten yang seharusnya UNDERPRICING berhasil diprediksi dengan benar sebagai UNDERPRICING.

C. Paired t-test

Untuk memastikan stabilitas dan generalisasi model, alur kerja pengembangan model dilakukan secara berulang sebanyak 30 kali menggunakan *seed* secara acak. Pendekatan ini dilakukan untuk memastikan bahwa hasil model yang diperoleh bukan terjadi secara kebetulan dan evaluasi antara 2 model yang dihasilkan memiliki perbedaan yang signifikan. Oleh karena itu, dilakukan *uji paired t-test* pada dua model terbaik. Berikut visualisasi data perbedaan akurasi *random forest* dan *neural network*:

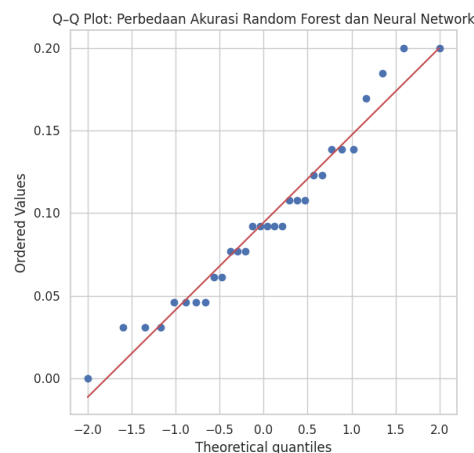


Gambar 6. Histogram perbedaan akurasi RF dan NN

Berdasarkan gambar 6 dan 7, dapat dilihat bahwa data selisih akurasi *random forest* dan *neural network* terdistribusi normal. Oleh karena itu, *paired t-test* dapat dilakukan.

TABEL V
PAIRED SAMPLES STATISTICS

	Column	N	Missing Count	Mean	Standard Deviation	Standard Error Mean
Pair 1	Random Forest	30	0	0,8497	0,0413	0,0075
Pair 1	Neural Network	30	0	0,7554	0,0432	0,0079



Gambar 7. QQ Plot perbedaan akurasi RF dan NN (Skewness = 0.426046
KS-test p-value = 0.77104)

Berikut adalah hasil *paired t-test* dua model terbaik pada penelitian ini.

TABEL VI
PAIRED SAMPLES TEST

t	df	p-value (2-tailed)	Mean	Standard Deviation	Standard Error Mean	CI (Lower Bound)	CI (Upper Bound)
9,9971	29	6,64E-11	0,0944	0,0517	0,0094	0,0751	0,1137

Berdasarkan tabel V dan tabel VI, dapat disimpulkan bahwa perbedaan akurasi *random forest* dan *neural network* adalah signifikan secara statistik. Hal ini ditunjukkan dengan nilai $t=9.9971$, $df=29$, $p<0.05$. *Random forest* menunjukkan rata-rata akurasi tertinggi (0.8497 ± 0.0075 , dari 30 iterasi). Hasil ini mengonfirmasi bahwa kinerja *random forest* bukan hasil kebetulan acak. Oleh karena itu, model *random forest* dapat digunakan untuk mengklasifikasi emiten yang *underpricing*.

D. Hasil Tahapan Deployment

Berikut adalah hasil klasifikasi model *random forest* menggunakan data emiten yang *listing* pada Januari 2025.

TABEL VII
HASIL DEPLOYMENT

kode_emiten	Prediction (is_underpricing?)	Prediction (is_underpricing?) (Confidence)
DGWG	UNDERPRICING	87%
KSIX	UNDERPRICING	72%
BRRC	UNDERPRICING	68%
YOII	UNDERPRICING	77%
RATU	UNDERPRICING	68%
HGII	UNDERPRICING	59%
CBDK	UNDERPRICING	66%
OBAT	UNDERPRICING	69%

Berdasarkan hasil klasifikasi pada tabel VII, 8 emiten yang *listing* pada Januari 2025 teridentifikasi *underpricing*. Hal ini menunjukkan bahwa harga saham emiten-emiten tersebut berpotensi mengalami kenaikan pada saat hari pertama *listing* di bursa. Informasi ini dapat digunakan oleh investor dalam pengambilan keputusan pembelian saham IPO tersebut dengan mempertimbangkan segala risiko yang ada.

IV. KESIMPULAN

Hasil penelitian menunjukkan bahwa algoritma *machine learning* dapat diterapkan untuk mengklasifikasi emiten yang *underpricing*. Skema dataset yang digunakan adalah dataset yang di-*oversampling* label minoritasnya dengan SMOTE dan dilakukan *split data* dengan perbandingan 80:20. Dari beberapa algoritma yang diterapkan, algoritma *random forest* menghasilkan tingkat akurasi tertinggi. Hasil evaluasi model *random forest* dengan *confusion matrix* menunjukkan tingkat akurasi sebesar 89,2%, F1-score 88.9%, dan AUC 0.946. Hasil uji *paired t-test* juga secara statistik mengonfirmasi bahwa hasil model *random forest* memiliki perbedaan signifikan dan unggul secara statistik. Model ini mengklasifikasikan 8 emiten yang IPO pada Januari 2025 termasuk *underpricing*.

DAFTAR PUSTAKA

- [1] Prof. J. Hartono, "Portofolio dan Analisis Investasi: Pendekatan Modul (Edisi 2)," 2022.
- [2] S. Husnan, "Dasar-Dasar Teori Portofolio dan Analisis Sekuritas, Edisi 3," Yogyakarta: UPPN STIM YKPN, 2019.
- [3] R. Sugiyanto, K. Indra Daniawan, W. Tangkas, F. Eko Mardiansyah, and J. Darmawan, "Analisis Underpricing Pada Perusahaan yang Melakukan IPO," *JIMEA / Jurnal Ilmiah MEA (Manajemen, Ekonomi, dan Akuntansi)*, vol. 7, no. 3, pp. 1517–1526, Dec. 2023, Accessed: Jul. 01, 2025. [Online]. Available: <https://www.journal.stiemb.ac.id/index.php/mea/article/view/3415>
- [4] M. Octafian, A. Wijayanti, and E. Masitoh, "Pengaruh DER, Roa, NPM dan EPS Terhadap Underpricing Studi Kasus: Perusahaan Yang Melakukan Initial Public Offering di BEI," *Jurnal Dinamika Ekonomi Pembangunan*, vol. 4, no. 1, pp. 15–20, 2021, doi: <https://doi.org/10.33005/jdep.v4i1.199>.
- [5] A. A. Taslim and I. Indrayenti, "Pengaruh Return on Equity, Financial Leverage dan Ukuran Perusahaan Terhadap Underpricing Saham," *EKONOMIKA45*, vol. 11, no. 1, pp. 411–422, Oct. 2023, Accessed: Jul. 01, 2025. [Online]. Available: <https://jurnaluniv45sby.ac.id/index.php/ekonomika/article/view/1875>
- [6] N. Kusuma Wardana, A. Suherman, and E. Eriswanto, "Pengaruh Return On Asset, Dan Debt Equity Ratio Terhadap Tingkat Underpricing," *Competitive Jurnal Akuntansi dan Keuangan*, vol. 5, no. 1, pp. 144–153, 2021, Accessed: Jul. 01, 2025. [Online]. Available: <https://jurnal.umt.ac.id/index.php/competitive/article/view/4776>
- [7] H. Alya Hafsa and Khairunnisa, "Tingkat Underpricing dan Faktor-Faktor yang Mempengaruhinya," *Jurnal Studi Akuntansi dan Keuangan*, vol. 6, no. 2, pp. 310–325, Dec. 2023, Accessed: Jul. 01, 2025. [Online]. Available: <https://akurasi.unram.ac.id/index.php/akurasi/article/view/406>
- [8] F. R. Maylani, J. Sudiman, and R. Heriyanto, "Pengaruh Faktor Financial dan Non Financial Terhadap Tingkat Underpricing Pada Perusahaan Sektor Consumer," *Jurnal Riset Akuntansi*, vol. 7, no. 1, pp. 191–102, Mar. 2024, Accessed: Jul. 01, 2025. [Online]. Available: <https://jra.politala.ac.id/index.php/JRA/article/view/304>
- [9] S. R. Apriliani and B. Bustamam, "Pengaruh Financial Leverage, Ukuran Perusahaan, dan Kepemilikan Institusional Terhadap Underpricing Saham Syariah Saat Initial Public Offering (IPO) di Bursa Efek Indonesia Pada Masa Pandemi COVID-19," *Jurnal Ilmiah Mahasiswa Ekonomi Akuntansi*, vol. 7, no. 1, pp. 118–130, 2022, doi: 10.24815/jimeka.v7i1.21062.
- [10] M. F. Rohman and I. R. Fitriati, "Pengaruh Umur Perusahaan, Ukuran Perusahaan, Return On Asset dan Debt to Equity Ratio Terhadap Underpricing Saham Pada Perusahaan Jasa Saat Initial Public Offering (IPO) Di Bursa Efek Indonesia (BEI) Periode 2020-2023," *EKOMA : Jurnal Ekonomi, Manajemen, Akuntansi*, vol. 3, pp. 2478–2488, Sep. 2024, Accessed: Jan. 07, 2025. [Online]. Available: <https://ulilbabbinstitute.id/index.php/EKOMA/article/view/5361>
- [11] S. Hakim and Fauzan, "Pengaruh Variabel Keuangan Dan Non Keuangan Terhadap Underpricing Saham Pada Saat Ipo Di Bursa Efek Indonesia (Periode 2020-2022)," *INNOVATIVE: Journal Of Social Science Research*, vol. 3, no. 5, pp. 2211–2223, 2023, Accessed: Jan. 07, 2025. [Online]. Available: <https://j-innovative.org/index.php/Innovative/article/view/4706>
- [12] R. N. R. Andriani, I. Hermansyah, and N. N. Wahid, "Pengaruh Current Ratio, Return On Asset dan Nilai Tukar Terhadap Underpricing," *JURNAL AKUNTANSI*, vol. 17, no. 1, pp. 70–80, 2023, doi: 10.37058/jak.v17i1.6740.
- [13] I. K. E. S. I Ketut Edi Suryawan and Gede Adi Yuniarta, "Pengaruh Financial Distress, Financial Leverage, Umur Perusahaan dan Inflasi Terhadap Underpricing Pada Perusahaan yang Melakukan Initial Public Offering (IPO)," *Vokasi : Jurnal Riset Akuntansi*, vol. 12, no. 3, pp. 148–159, 2023, doi: 10.23887/vjra.v12i3.69101.
- [14] Sonia, F. Soejono, A. S. Mendari, and M. Vinice, "Pengaruh Leverage, Profitabilitas dan Umur Perusahaan Terhadap Tingkat Underpricing pada Perusahaan yang Melakukan IPO di BEI Tahun 2020-2023," *Jurnal Ilmiah Ekonomi Dan Bisnis*, vol. 14, pp. 21–34, Mar. 2025, Accessed: May 07, 2025. [Online]. Available: <https://jurnal.mdp.ac.id/index.php/forbiswira/article/view/9752>
- [15] M. W. Putri and D. Ekaningtiyas, "Pengaruh Financial Leverage, Profitabilitas, dan Ukuran Perusahaan terhadap Tingkat Underpricing Saham Initial Public Offering di Bursa Efek Indonesia Periode 2020-2023," *IDEI: Jurnal Ekonomi & Bisnis*, vol. 5, no. 2, pp. 96–109, Dec. 2024, doi: 10.38076/idejeb.v5i2.295.
- [16] R. T. Yunarni and N. Z. Dina, "Data Mining dan Machine Learning Dengan Orange3 Tutorial dan Aplikasinya," 2022.
- [17] I. Kurniawan, N. Rahaningsih, and T. Suprpti, "Implementasi Algoritma Regresi Linier dan K-Nearest Neighbor Untuk Prediksi Harga Rumah," 2024. Accessed: May 08, 2025. [Online]. Available: <https://ejournal.itn.ac.id/index.php/jati/article/view/8402>
- [18] S. Mulia Agustina and D. Gustian, "Analisis Kinerja Algoritma Prediksi Saham pada PT GoTo Gojek Tokopedia Tbk (GOTO)," vol. 11, no. 1, pp. 53–67, 2025, Accessed: May 08, 2025. [Online]. Available: <https://rekayasa.nusaputra.ac.id/index>
- [19] R. M. A. Ananda and R. Kurniawan, "Perbandingan Algoritma Regresi Linier Dengan Neural Network Untuk Memprediksi Harga Saham Bank JAGO," 2024. Accessed: May 08, 2025. [Online]. Available: <https://ejournal.itn.ac.id/index.php/jati/article/view/8998>
- [20] S. Amaliah, M. Nusrang, and A. Aswi, "Penerapan Metode Random Forest Untuk Klasifikasi Varian Minuman Kopi di Kedai Kopi Konijawa Bantaeng," *VARIANSI: Journal of Statistics and Its application on Teaching and Research*, vol. 4, no. 3, pp. 121–127, Dec. 2022, doi: 10.35580/variansiunm31.
- [21] A. C. Muhammad *et al.*, *Dasar-dasar Pembelajaran Mesin*. PT Sada Kurnia Pustaka, 2023.