

Prediksi Pergerakan Harga Saham Berbasis Rasio Keuangan Menggunakan Metode Artificial Neural Network (ANN)

1st Nakhwa Azizah
Fakultas Informatika
Telkom University
Bandung, Indonesia
nakhwaazizah@student.telkomuniversity.ac.id

2nd Deni Saepudin
Fakultas Informatika
Telkom University
Bandung, Indonesia
denisaepudin@telkomuniversity.ac.id

Abstrak — Prediksi pergerakan harga saham merupakan tantangan di dunia investasi karena sifatnya fluktuatif dan dipengaruhi berbagai faktor. Penelitian ini menggunakan *Artificial Neural Network* (ANN) untuk memprediksi pergerakan harga saham mingguan pada 19 perusahaan indeks LQ45. Data mencakup harga penutupan mingguan (*Close*) dan enam indikator rasio keuangan: EPS, ROA, ROE, P/B Value, P/E Ratio, dan DER. Pergerakan harga saham diklasifikasikan menjadi 1 (naik), 0 (stagnan), dan -1 (turun), berdasarkan tiga *threshold* perubahan harga (1%, 2%, 3%). Penelitian ini menerapkan *Principal Component Analysis* (PCA) untuk mengurangi *noise* dan meningkatkan performa model. Hasil menunjukkan *threshold* optimal adalah 2%, dengan akurasi dan F1-Score yang lebih baik. Model berbasis data historis menunjukkan performa terbaik dalam memprediksi pergerakan harga saham, dengan rata-rata akurasi mencapai 0,92. Sebaliknya, model yang mengintegrasikan rasio keuangan tanpa PCA memiliki akurasi lebih rendah sebesar 0,5 hingga 0,7. Hal ini menunjukkan rasio keuangan memiliki korelasi rendah terhadap pergerakan harga saham dalam jangka pendek. Penerapan PCA pada data historis dan rasio keuangan meningkatkan akurasi hingga setara dengan model berbasis data historis saja. Uji statistik menunjukkan bahwa tidak terdapat perbedaan signifikan dalam akurasi dan varians antara model berbasis data historis dan model kombinasi dengan PCA.

Kata kunci— prediksi pergerakan harga saham, *artificial neural network*, data historis, rasio keuangan, *threshold*, PCA.

I. PENDAHULUAN

Dalam penelitian ini dilakukan pembahasan terkait dengan masalah bagaimana memprediksi pergerakan harga saham dengan memanfaatkan data rasio keuangan sebagai indikator. Masalah ini penting karena pergerakan harga saham mencerminkan perubahan nilai saham dan sering digunakan untuk menggambarkan kinerja pasar [1]. Rasio keuangan dipilih sebagai indikator karena dapat menunjukkan kesehatan keuangan perusahaan dan prospek pertumbuhannya di masa depan [2]. Dengan menggunakan data historis harga penutupan mingguan (*Close*) dan indikator rasio keuangan, model prediktif dapat dikembangkan untuk memprediksi pergerakan harga saham [2]. Salah satu metode yang efektif untuk membangun model ini adalah *Artificial Neural Network* (ANN), karena ANN bekerja dengan meniru cara otak manusia dalam memproses informasi dan mengenali pola kompleks dalam data [3][4].

Penelitian sebelumnya telah menunjukkan keunggulan ANN dalam prediksi harga saham. Penelitian Patalay dan Bandlamudi [2] menunjukkan bahwa ANN berbasis data fundamental dapat memprediksi harga saham jangka panjang dengan metode regresi, yang menghasilkan akurasi *Root Relative Squared Error* (RRSE) sebesar 8.877%. Penelitian Muhammad Ali et al. [1] membandingkan ANN dan *Support Vector Machine* (SVM) dalam memprediksi pergerakan harga penutupan harian dengan metode klasifikasi biner (naik atau turun). Hasilnya menunjukkan bahwa ANN lebih unggul dalam akurasi dan F1-Score. Selain itu, penelitian Parshv

Chhajer et al. [5] membuktikan kemampuan ANN untuk memprediksi arah pergerakan harga saham dalam memodelkan data non-linear dan memberikan hasil prediksi yang baik, terutama pada data dengan volatilitas tinggi.

Namun, sebagian besar penelitian sebelumnya hanya menggunakan satu jenis data, seperti data historis atau data fundamental seperti rasio keuangan. Pendekatan ini menyisakan peluang untuk memanfaatkan potensi dari integrasi kedua jenis data. Oleh karena itu, penelitian ini bertujuan untuk mengembangkan model prediktif pergerakan harga saham berbasis rasio keuangan menggunakan metode ANN.

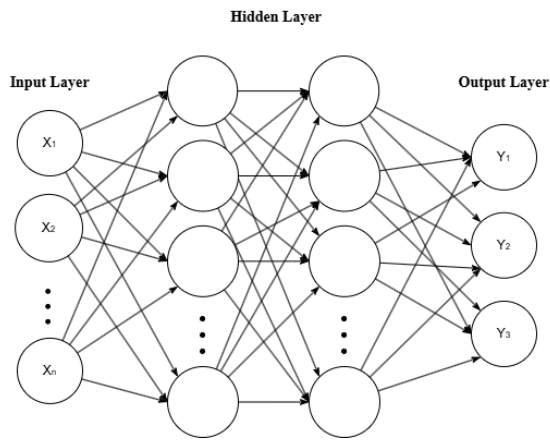
Data dalam penelitian ini diklasifikasikan ke dalam tiga kategori yaitu 1 (naik), 0 (stagnan), dan -1 (turun). Kategori ini didasarkan pada persentase perubahan harga penutupan mingguan dengan tiga *threshold* yaitu 1%, 2%, dan 3%. Tujuannya adalah untuk menyederhanakan pengelompokan pergerakan harga saham dan menangkap pola yang signifikan. Indikator rasio keuangan seperti EPS (*Earnings Per Share*), ROA (*Return on Assets*), ROE (*Return on Equity*), P/B Value (*Price-to-Book Value*), P/E Ratio (*Price-to-Earnings Ratio*), dan DER (*Debt-to-Equity Ratio*) digunakan untuk mendukung prediksi pergerakan harga saham. Selain itu, teknik *Principal Component Analysis* (PCA) diterapkan untuk mengurangi *noise* pada data rasio keuangan.

Pendekatan ini diharapkan dapat meningkatkan akurasi prediksi pergerakan harga saham serta memberikan analisis mendalam tentang hubungan rasio keuangan dengan pergerakan harga saham. Penelitian ini juga diharapkan berkontribusi dalam pengembangan metode prediktif yang andal dan aplikatif di dunia investasi.

II. KAJIAN TEORI

A. *Artificial Neural Network*

Artificial Neural Network (ANN) adalah algoritma pembelajaran mesin yang terinspirasi dari cara kerja sistem saraf manusia. ANN sering digunakan untuk tugas klasifikasi, seperti prediksi harga saham [6][7]. Salah satu jenis ANN yang sering digunakan adalah *Multilayer Perceptron* (MLP), yang terdiri dari tiga lapisan utama, yaitu lapisan input, tersembunyi, dan output. MLP dirancang untuk menangkap hubungan non-linier dalam data [8]. Gambar 2.1 menunjukkan arsitektur MLP.



GAMBAR 2.1
Arsitektur Multilayer Perceptron

Pada arsitektur ini, lapisan input menerima data dalam bentuk vektor fitur, sesuai dengan jumlah fitur pada dataset. Lapisan tersembunyi mengekstraksi pola dari data masukan dengan kombinasi linear. Kombinasi linear ini kemudian diproses menggunakan fungsi aktivasi, seperti ReLU, untuk memperkenalkan non-linearitas pada model. Hasil dari lapisan tersembunyi diteruskan ke lapisan output untuk menghasilkan prediksi akhir [6]. Kombinasi linear pada setiap neuron dapat dirumuskan sebagai berikut:

$$z = \sum_{i=1}^n w_i \cdot x_i + b \quad (2.1)$$

Keterangan:

- z : Hasil kombinasi linear dalam *neuron*
- n : Jumlah total input ke *neuron*
- w_i : Bobot untuk input ke- i
- x_i : Nilai input ke- i
- b : Bias

Proses pelatihan ANN melibatkan dua tahap utama, yaitu forward propagation dan backward propagation. Prediksi dibandingkan dengan nilai target sebenarnya menggunakan fungsi loss untuk menghitung error [9]. Fungsi loss yang digunakan adalah Categorical Cross-Entropy, yang dirumuskan sebagai berikut:

$$L_{(y,\hat{y})} = \sum_{i=1}^N y_i \cdot \log(\hat{y}_i) \quad (2.2)$$

Keterangan:

- $L_{(y,\hat{y})}$: Nilai loss Categorical Cross-Entropy
- y_i : Nilai sebenarnya dari label target untuk kelas ke- i
- \hat{y}_i : Probabilitas prediksi oleh model untuk kelas ke- i
- N : Jumlah total kelas

Categorical Cross-Entropy digunakan untuk tugas klasifikasi multi-kelas karena efektif dalam mengevaluasi seberapa baik model memprediksi probabilitas kelas yang benar [10]. Error yang dihasilkan akan digunakan dalam backward propagation untuk menghitung bobot melalui algoritma optimisasi. Proses pembaruan dilakukan dengan menghitung gradien error melalui algoritma optimisasi [9]. Dalam penelitian ini, algoritma optimisasi Adam digunakan

untuk memperbarui bobot. Adam dipilih kemampuannya menangani gradien yang fluktuatif.

Categorical Cross-Entropy digunakan untuk tugas klasifikasi multi-kelas karena efektif dalam mengevaluasi seberapa baik model memprediksi probabilitas kelas yang benar [10]. Error yang dihasilkan akan digunakan dalam backward propagation untuk menghitung bobot melalui algoritma optimisasi. Proses pembaruan dilakukan dengan menghitung gradien error melalui algoritma optimisasi [9]. Dalam penelitian ini, algoritma optimisasi Adam digunakan untuk memperbarui bobot. Adam dipilih kemampuannya menangani gradien yang fluktuatif.

Model ANN pada penelitian ini dirancang untuk tugas klasifikasi multi-kelas, khususnya prediksi pergerakan harga saham. Untuk meningkatkan akurasi dan kemampuan generalisasi model, beberapa teknik diterapkan. Dengan pendekatan ini, model diharapkan memberikan prediksi yang akurat dan menangkap pola pergerakan harga saham dengan baik.

B. Return Saham

Return saham adalah tingkat keuntungan atau kerugian dari investasi saham dalam periode tertentu [11]. Dalam penelitian ini, *return* saham dihitung berdasarkan perubahan persentase harga penutupan mingguan. Rumus adalah:

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}} \times 100\% \quad (2.3)$$

Keterangan:

- R_t : Nilai *return* pada waktu t dalam bentuk persentase
- P_t : Harga penutupan saham pada waktu t
- P_{t-1} : Harga penutupan saham sebelum waktu t

Return saham menjadi parameter penting dalam penelitian ini karena digunakan untuk mengklasifikasikan pergerakan harga saham ke dalam kategori naik, stagnan, dan turun.

C. Korelasi Spearman

Korelasi Spearman atau *spearman's rank correlation coefficient* adalah metode statistik non-parametrik untuk mengukur hubungan monotonic antara dua variabel [12]. Korelasi Spearman sering digunakan untuk data ordinal atau data interval yang tidak memenuhi asumsi distribusi normal maupun hubungan linier, menjadikannya alternatif yang fleksibel dibandingkan korelasi Pearson [13]. Rumus menghitung korelasi Spearman adalah [14]:

$$\rho = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)} \quad (2.4)$$

Keterangan:

- ρ : Koefisien korelasi Spearman
- d_i : Selisih peringkat antara dua variabel untuk pengamatan ke- i
- n : Jumlah total pengamatan atau pasangan data

Dalam penelitian ini, Korelasi Spearman digunakan untuk menganalisis hubungan antara fitur rasio keuangan, seperti EPS, ROA, ROE, P/B Value, P/E Ratio, dan DER, terhadap klasifikasi pergerakan harga saham. Penggunaan Korelasi Spearman dianggap tepat karena data klasifikasi pergerakan harga saham bersifat ordinal.

D. Principal Component Analysis (PCA)

Principal Component Analysis (PCA) adalah metode reduksi dimensi yang mengubah data kompleks menjadi

sekumpulan variabel yang tidak saling berkorelasi, yang disebut komponen utama, tanpa menghilangkan informasi penting. Komponen utama bekerja dengan menangkap variansi maksimum sehingga memungkinkan analisis dan interpretasi menjadi lebih sederhana [15][16]. Dalam penelitian ini, PCA diterapkan pada kombinasi semua fitur rasio keuangan dengan menggunakan empat komponen utama, yang berhasil menjelaskan variansi kumulatif hingga sekitar 90%.

E. Pengujian Hipotesis Statistik

Pengujian hipotesis statistik adalah proses untuk membuat keputusan berdasarkan data sampel tentang suatu pernyataan atau dugaan yang berkaitan dengan karakteristik populasi [17]. Pengujian hipotesis ini bertujuan untuk menentukan apakah data sampel mendukung penolakan hipotesis nol atau tidak, dengan memperhatikan tingkat signifikansi tertentu (α).

Uji t digunakan untuk membandingkan rata-rata satu atau lebih kelompok untuk menentukan apakah terdapat perbedaan signifikan secara statistik [18]. Rumus uji t untuk dua sampel adalah sebagai berikut:

$$t_{hitung} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\left(\frac{s_1^2}{n_1}\right) + \left(\frac{s_2^2}{n_2}\right)}} \quad (2.5)$$

Derajat kebebasan (df) dihitung dengan rumus:

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\left[\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1}\right] + \left[\frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}\right]} \quad (2.6)$$

Keterangan:

\bar{x}_1, \bar{x}_2 : Rata-rata dari sampel 1 dan 2

s_1^2, s_2^2 : Selisih peringkat antara dua variabel untuk pengamatan ke- i

n_1, n_2 : Jumlah sampel 1 dan 2

Uji ini digunakan untuk mengevaluasi rata-rata akurasi model berbasis data historis dengan model PCA dalam penelitian ini untuk menentukan apakah terdapat perbedaan signifikan.

Uji F digunakan untuk membandingkan variansi dua atau lebih kelompok guna menguji hipotesis terkait kesetaraan variansi [18]. Rumus uji F adalah sebagai berikut:

$$F_{hitung} = \frac{s_1^2}{s_2^2} \quad (2.7)$$

Derajat kebebasan untuk uji F dihitung sebagai berikut:

$$df_1 = n_1 - 1 \quad (2.8)$$

$$df_2 = n_2 - 1 \quad (2.9)$$

Keterangan:

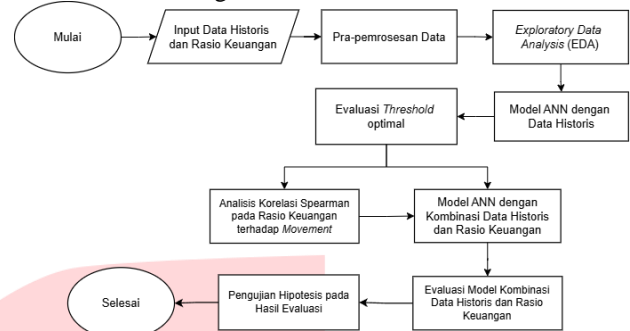
s_1^2, s_2^2 : Selisih peringkat antara dua variabel untuk pengamatan ke- i

n_1, n_2 : Jumlah sampel 1 dan 2

Uji ini digunakan dalam penelitian ini untuk membandingkan variansi model berbasis data historis dengan model PCA untuk mengevaluasi stabilitas akurasi masing-masing model.

III. METODE

1. Desain Perancangan Sistem



GAMBAR 3.1

Desain Perancangan Sistem

Gambar 3.1 menunjukkan desain perancangan sistem secara keseluruhan dalam penelitian. Desain sistem mencakup beberapa tahapan utama, yaitu pengumpulan data historis dan rasio keuangan, pra-pemrosesan data, *exploratory data analysis* (EDA), evaluasi *threshold* optimal, analisis korelasi menggunakan Spearman pada rasio keuangan terhadap *movement*, pembangunan model ANN dengan data historis maupun kombinasi data historis dan rasio keuangan, evaluasi performa model, serta pengujian hipotesis pada hasil evaluasi.

2. Data

Data yang digunakan dalam penelitian ini mencakup data historis saham diperoleh dari Yahoo Finance dalam bentuk mingguan dan data rasio keuangan diperoleh dari Stockbit dalam bentuk kuartal. Penelitian ini menggunakan data dari 19 perusahaan yang terdaftar dalam indeks LQ45 Bursa Efek Indonesia (BEI). Tabel 3.1 menunjukkan daftar perusahaan LQ45 yang digunakan dalam penelitian.

TABEL 3.1
Daftar Perusahaan LQ45

No.	Sektor	Kode Saham	Nama Perusahaan
1.	Energi	ADRO	PT Adaro Energy Tbk
2.		AKRA	PT Akasha Wira International Tbk
3.		ITMG	PT Indo Tambangraya Megah Tbk
4.		PGAS	PT Perusahaan Gas Negara Tbk
5.		PTBA	PT Bukit Asam Tbk
6.	Keuangan	BBCA	PT Bank Central Asia Tbk
7.		BBNI	PT Bank Negara Indonesia (Persero) Tbk
8.		BBRI	PT Bank Rakyat Indonesia (Persero) Tbk

9.		BBTN	PT Bank Tabungan Negara (Persero) Tbk
10.		BMRI	PT Bank Mandiri (Persero) Tbk
11.	Barang Baku	ANTM	PT Aneka Tambang Tbk
12.		INCO	PT Vale Indonesia Tbk
13.		INKP	PT Indah Kiat Pulp & Paper Tbk
14.		INTP	PT Indocement Tunggak Prakarsa Tbk
15.		SMGR	PT Semen Indonesia (Persero) Tbk
16.	Infrastruktur	EXCL	PT XL Axiata Tbk
17.		ISAT	PT Indosat Ooredoo Hutchison Tbk
18.		JSMR	PT Jasa Marga (Persero) Tbk
19.		TLKM	PT Telkom Indonesia (Persero) Tbk

Data historis dan rasio keuangan yang digunakan dalam penelitian ini adalah sebagai berikut:

1. *Close*

Harga penutupan saham atau nilai transaksi terakhir tercatat pada akhir sesi perdagangan suatu periode.

2. *EPS (Earnings Per Share)*

Metrik keuangan yang mengukur profitabilitas setiap saham dan mencerminkan kemampuan perusahaan menghasilkan keuntungan [19].

3. *ROA (Return on Assets)*

Menunjukkan efektivitas perusahaan dalam memanfaatkan asetnya untuk menghasilkan laba [19].

4. *ROE (Return on Equity)*

Metrik keuangan yang mengukur profitabilitas perusahaan yang dihasilkan dari ekuitas pemegang saham [19].

5. *P/B Value (Price-to-Book Value)*

Membandingkan nilai pasar ekuitas perusahaan dengan nilai bukunya dan membantu investor menilai apakah suatu saham dihargai terlalu rendah atau terlalu tinggi di pasar [20].

6. *P/E Ratio (Price-to-Earnings Ratio)*

Membandingkan harga saham perusahaan dengan laba per saham (EPS) untuk menilai valuasi saham dan mengukur ekspektasi pasar terhadap kinerja perusahaan [21].

7. *DER (Debt-to-Equity Ratio)*

Mengukur perbandingan antara total utang perusahaan dengan ekuitas pemegang sahamnya sebagai sumber pendanaan perusahaan [22].

3. Pra-pemrosesan Data

Pra-pemrosesan data dilakukan untuk memastikan data siap digunakan dalam pemodelan. Langkah-langkah yang dilakukan meliputi pembersihan data, normalisasi data,

perhitungan return saham, dan klasifikasi pergerakan harga saham. Adapun klasifikasi ini menggunakan tiga skenario *threshold*:

a. *Threshold 1%*

$$Klasifikasi = \begin{cases} 1, & \text{jika } R_t > 1 \\ 0, & \text{jika } -1 \leq R_t \leq 1 \\ -1, & \text{jika } R_t < -1 \end{cases}$$

b. *Threshold 2%*

$$Klasifikasi = \begin{cases} 1, & \text{jika } R_t > 2 \\ 0, & \text{jika } -2 \leq R_t \leq 2 \\ -1, & \text{jika } R_t < -2 \end{cases}$$

c. *Threshold 3%*

$$Klasifikasi = \begin{cases} 1, & \text{jika } R_t > 3 \\ 0, & \text{jika } -3 \leq R_t \leq 3 \\ -1, & \text{jika } R_t < -3 \end{cases}$$

1. Transformasi Data ke Bentuk Numerik

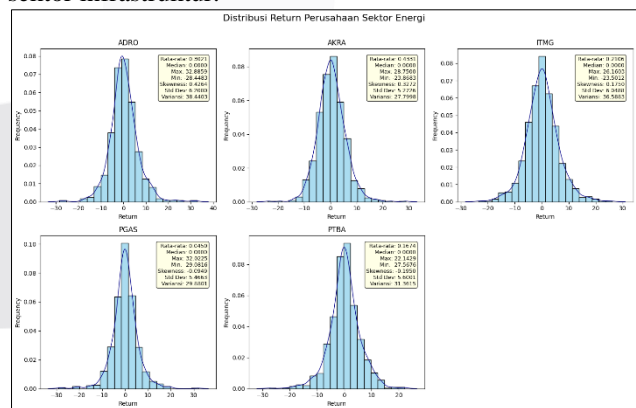
Variabel movement dikonversi ke bentuk numerik menggunakan One-Hot Encoding. Langkah ini dilakukan karena model ANN hanya dapat menerima input dalam format numerik.

2. Pembagian Data

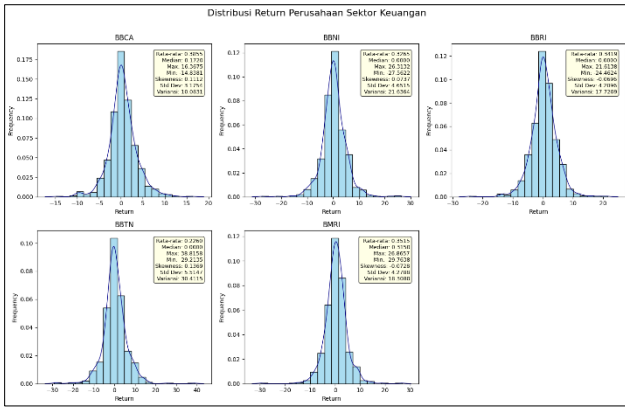
Dataset dibagi secara terurut untuk menjaga sifat time series. Sebanyak 80% data digunakan untuk pelatihan (training), dan 20% sisanya digunakan untuk pengujian (testing).

4. *Exploratory Data Analysis (EDA)*

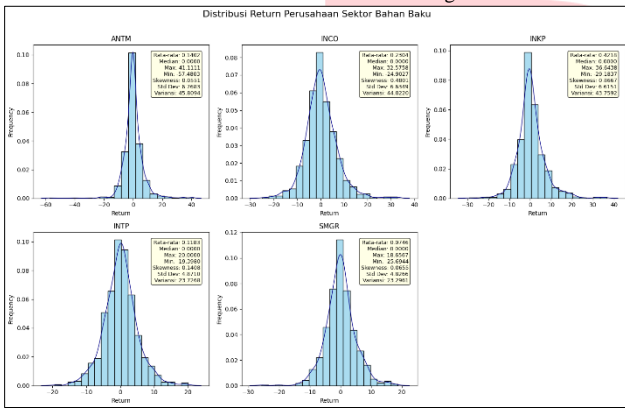
EDA dilakukan dengan menggunakan histogram untuk membantu menggambarkan distribusi *return* mingguan setiap perusahaan. Visualisasi dilengkapi dengan statistik deskriptif seperti rata-rata, median, standar deviasi, dan skewness. Gambar 3.2 menunjukkan distribusi *return* sektor energi, Gambar 3.3 menunjukkan distribusi *return* sektor keuangan, Gambar 3.4 menunjukkan distribusi *return* sektor bahan bahu, dan Gambar 3.5 menunjukkan distribusi *return* sektor infrastruktur.



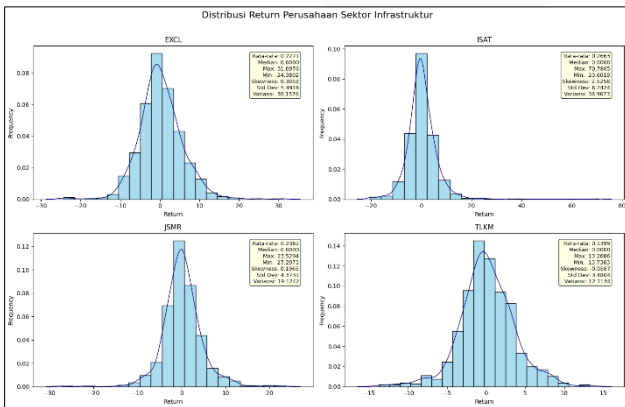
GAMBAR 3.2
Distribusi *Return* Sektor Energi



GAMBAR 3.3
Distribusi Return Sektor Keuangan



GAMBAR 3.4
Distribusi Return Sektor Bahan Baku



GAMBAR 3.5
Distribusi Return Infrastruktur

Berdasarkan hasil EDA, rata-rata *return* sebagian besar perusahaan di berbagai sektor berada pada kisaran positif. Sebaran data menunjukkan bahwa sektor energi memiliki persebaran *return* terbesar, sedangkan sektor infrastruktur memiliki persebaran *return* terkecil. Distribusi *return* diseluruh sektor cenderung simetris dengan nilai *skewness* yang rendah, mengindikasikan peluang yang seimbang antara *return* positif dan negatif. Hasil ini memberikan wawasan awal untuk mengidentifikasi pola penyebaran dan tren *return* saham di setiap sektor.

5. Perancangan Model Prediksi

Model ANN dirancang untuk menangani klasifikasi multi-kelas. Tabel 3.2 menyajikan daftar arsitektur dan parameter yang digunakan dalam model ANN untuk mendukung proses klasifikasi tersebut.

TABEL 3.2

Arsitektur dan Parameter Model

Komponen	Detail
<i>Input layers</i>	Jumlah fitur dalam data
<i>Hidden layer 1</i>	64 <i>neurons</i> , ReLU
<i>Hidden layer 2</i>	32 <i>neurons</i> , ReLU
<i>Output layer</i>	3 <i>neurons</i> , <i>Softmax</i>
<i>Regularization</i>	<i>Kernel Regularizer L2</i>
<i>Dropout</i>	30% (<i>hidden layers</i>)
<i>Stopping criteria (epochs)</i>	<i>Patience</i> = 10 (<i>EarlyStopping</i>)
<i>Optimizer</i>	Adam
<i>Loss function</i>	<i>Categorical Cross-entropy</i>
<i>Metrics</i>	Akurasi
<i>Epochs</i>	50
<i>Batch size</i>	16

Input layer menyesuaikan jumlah fitur pada data yang telah diproses. *Hidden layers* menggunakan fungsi aktivasi ReLU karena kemampuannya menangani non-linearitas. *Dropout* sebesar 30% diterapkan untuk mencegah *overfitting*, sedangkan *kernel regularizer L2* membantu menjaga stabilitas bobot model. Optimisasi Adam dipilih karena efisiensinya dalam menyesuaikan *learning rate* selama pelatihan, dan fungsi *loss Categorical Cross-entropy* digunakan untuk menghitung kesalahan pada klasifikasi multi-kelas secara akurat.

6. Evaluasi Performansi Model

a. Confusion Matrix

Confusion Matrix berisi informasi tentang jumlah prediksi benar dan salah. Tabel 3.3 menunjukkan tabel *Confusion Matrix*.

TABEL 3.3
Confusion Matrix

		Prediksi	
		True	False
Aktual	True	TP	FP
	False	FN	TN

b. Metrik Evaluasi

Berdasarkan *Confusion Matrix*, beberapa metrik evaluasi digunakan untuk mengukur performansi model adalah sebagai berikut:

1. Akurasi

Akurasi mengukur persentase prediksi yang benar dari total prediksi yang dilakukan oleh model. Akurasi dihitung menggunakan rumus:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (3.1)$$

2. Precision

Precision mengukur keakuratan prediksi kelas positif. *Precision* dihitung menggunakan rumus:

$$Precision = \frac{TP}{TP + FP} \quad (3.2)$$

3. Recall

Recall mengukur sensitivitas model untuk mendeteksi kelas positif. *Recall* dihitung menggunakan rumus:

$$Recall = \frac{TP}{TP + FN} \quad (3.3)$$

4. F1-Score

F1-Score adalah rata-rata harmonis antara presisi dan *recall*. F1-Score dihitung dengan rumus:

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (3.4)$$

Evaluasi model dilakukan menggunakan akurasi, *precision*, *recall*, dan F1-Score untuk setiap kombinasi data dan *threshold*. Penggunaan keempat metrik ini memberikan analisis menyeluruh terhadap performa model.

IV. HASIL DAN PEMBAHASAN

1. Percobaan Mencari *Threshold* Optimal

Tahap pertama percobaan adalah menentukan nilai *threshold* optimal untuk mengklasifikasikan pergerakan harga saham (*movement*). *Threshold* yang diuji yaitu 1%, 2%, dan 3% yang diterapkan pada data *return* mingguan. Percobaan dilakukan dengan membangun model ANN untuk setiap *threshold*, yang kemudian diuji performanya menggunakan data *testing*.

Penentuan *threshold* sangat penting dilakukan di awal percobaan karena nilai *threshold* akan memengaruhi bagaimana model mengklasifikasikan pergerakan harga saham. Dengan menentukan *threshold* di awal, kita dapat menyesuaikan sensitivitas model terhadap fluktuasi harga saham dan mengidentifikasi kategori yang relevan sesuai dengan karakteristik pasar.

2. Percobaan Mencari Korelasi Fitur Terkuat

Setelah *threshold* optimal ditentukan, langkah selanjutnya adalah mengidentifikasi tingkat hubungan antara variabel rasio keuangan dengan pergerakan harga saham (*movement*). Hubungan tersebut dihitung menggunakan korelasi Spearman, dan hasilnya divisualisasikan dalam bentuk *barchart*. Berdasarkan hasil tersebut, variabel rasio keuangan dengan nilai korelasi tertinggi dipilih untuk digunakan dalam kombinasi fitur pada tahap selanjutnya.

3. Percobaan Kombinasi Fitur

Berdasarkan analisis sebelumnya, percobaan dilakukan dengan menguji berbagai kombinasi fitur untuk menilai kontribusi masing-masing variabel terhadap kemampuan model dalam memprediksi *movement*. Sebanyak sepuluh kombinasi fitur diuji, termasuk kombinasi seluruh variabel rasio keuangan (*all fundamental*) pada percobaan terakhir. Berikut daftar kombinasi fitur yang diuji::

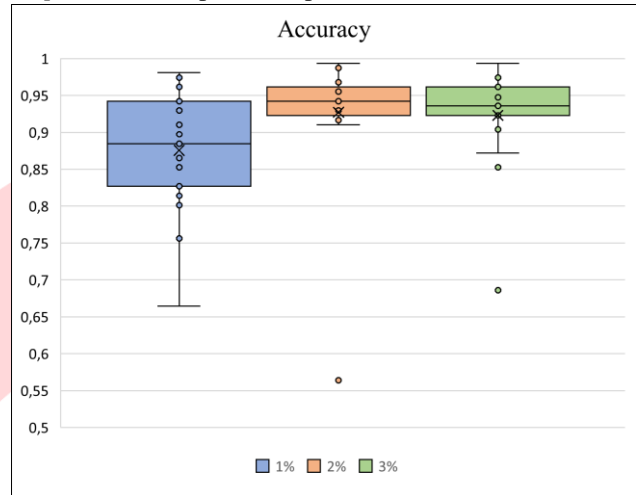
1. Historis
2. Historis + P/B Value
3. Historis + ROA
4. Historis + ROE
5. Historis + P/B Value + ROA
6. Historis + P/B Value + ROE
7. Historis + ROA + ROE
8. Historis + P/B Value + ROA + ROE
9. Historis + *All Fundamental*

10. Historis + *All Fundamental* + PCA

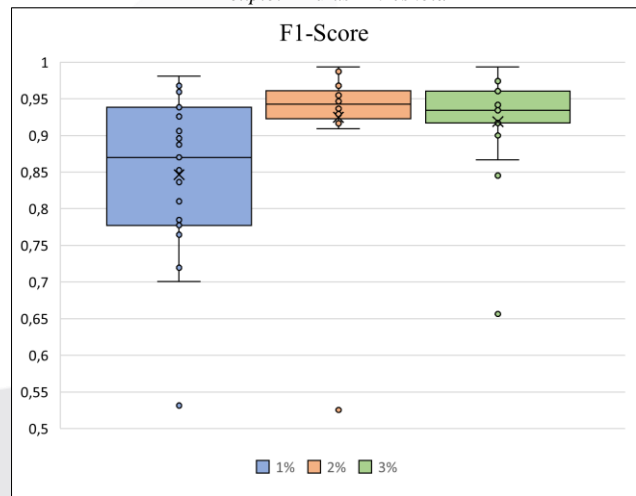
Pendekatan ini memastikan bahwa variabel rasio keuangan dengan korelasi kuat digunakan secara optimal, sementara kombinasi fitur lainnya memberikan gambaran tambahan mengenai kontribusi variabel-variabel tersebut terhadap prediksi pergerakan harga saham.

4. Hasil Penentuan *Threshold* Optimal

Gambar 4.1 menunjukkan *boxplot* akurasi pada setiap *threshold* yang diuji, sementara Gambar 4.2 menampilkan *boxplot* F1-Score pada setiap *threshold*.



GAMBAR 4.1
Boxplot Akurasi Threshold

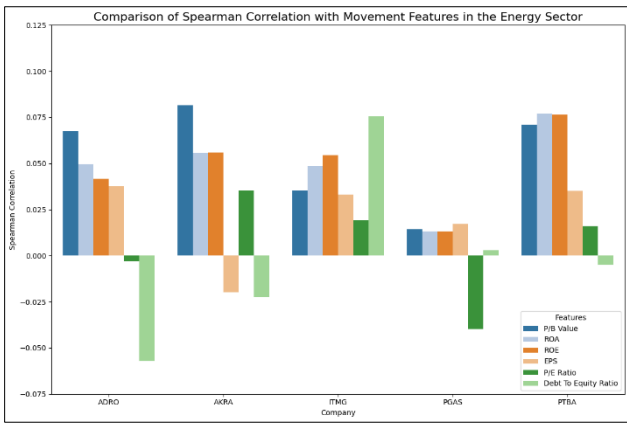


GAMBAR 4.2
Boxplot F1-Score Threshold

Dari Gambar 4.1 dan Gambar 4.2, dapat dilihat bahwa *threshold* 1% menghasilkan persebaran akurasi dan *F1-Score* yang paling bervariasi. Rentang nilai akurasi berada antara 0,7–0,95, dengan beberapa *outlier* pada nilai *F1-Score* di bawah 0,6. *Threshold* 2% memberikan hasil terbaik, dengan nilai rata-rata akurasi dan *F1-Score* yang lebih tinggi dibandingkan *threshold* lainnya yaitu sekitar 0,9.

5. Hasil Korelasi Fitur Terkuat

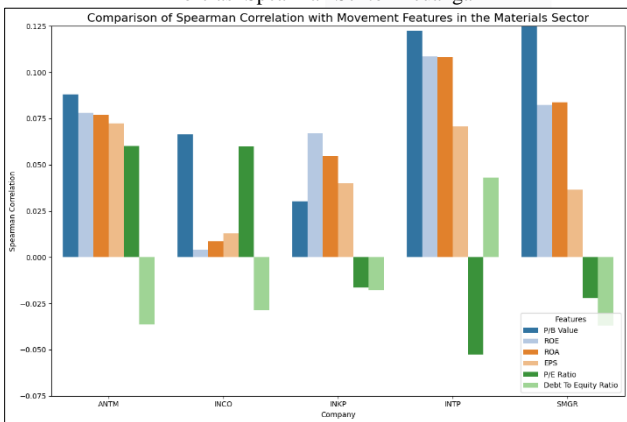
Gambar 4.3 menunjukkan hasil korelasi fitur pada sektor energi, Gambar 4.4 menunjukkan hasil korelasi fitur pada sektor keuangan, Gambar 4.5 menunjukkan hasil korelasi fitur pada sektor bahan baku, Gambar 4.6 menunjukkan hasil korelasi fitur pada sektor infrastruktur.



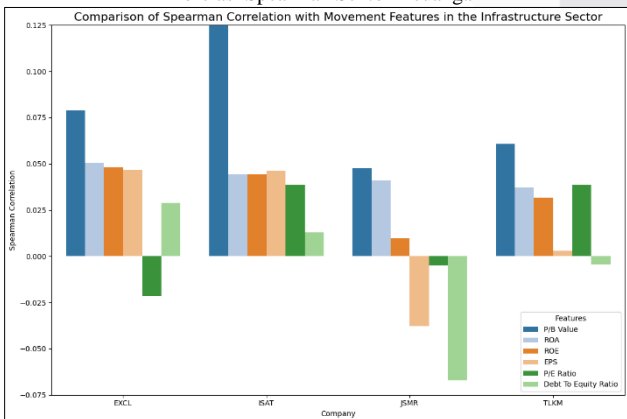
GAMBAR 4.3
Korelasi Spearman Sektor Energi



GAMBAR 4.4
Korelasi Spearman Sektor Keuangan



GAMBAR 4.5
Korelasi Spearman Sektor Keuangan



GAMBAR 4.6
Korelasi Spearman Sektor Infrastruktur

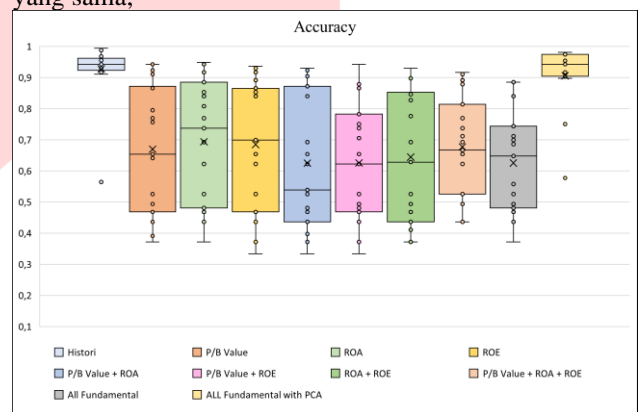
Dari hasil korelasi yang ditunjukkan pada Gambar 4.3 hingga Gambar 4.6, dapat disimpulkan beberapa hal sebagai berikut:

1. P/B Value menunjukkan korelasi positif yang konsisten di semua sektor.
2. ROA menunjukkan korelasi yang kuat di sektor bahan baku dan energi, sementara ROE relevan pada sektor bahan baku dan keuangan
3. DER secara umum menunjukkan korelasi negatif, yang mengindikasikan rasio ini kurang relevan dalam memprediksi pergerakan harga saham.

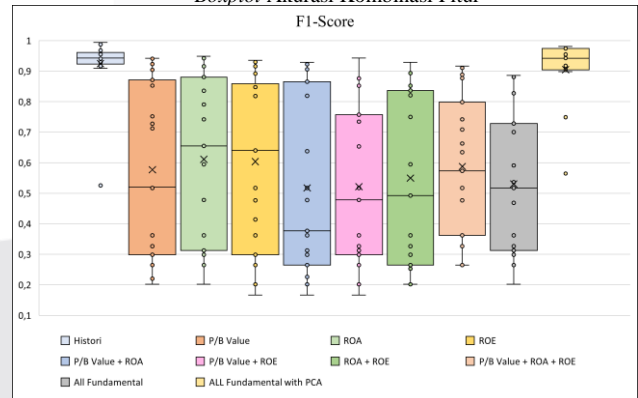
P/B Value, ROA, dan ROE dipilih sebagai variabel utama untuk dievaluasi lebih lanjut dalam eksperimen kombinasi fitur.

6. Hasil Kombinasi Fitur

Gambar 4.7 menunjukkan *boxplot* nilai akurasi untuk setiap kombinasi fitur yang diuji, sedangkan Gambar 4.8 menunjukkan *boxplot* nilai F1-Score untuk kombinasi fitur yang sama,



GAMBAR 4.7
Boxplot Akurasi Kombinasi Fitur



GAMBAR 4.8
Boxplot F1-Score Kombinasi Fitur

Dari hasil kedua *boxplot*, model yang menggunakan data historis saja memiliki rata-rata akurasi dan F1-Score mencapai 0,92. Sebaliknya, penambahan variabel fundamental seperti P/B Value, ROA, dan ROE, baik secara individu maupun kombinasi, justru menurunkan rata-rata akurasi dan F1-Score menjadi sekitar 0,5 hingga 0,7. Penggunaan PCA dengan empat komponen pada kombinasi semua fitur memberikan sedikit peningkatan performa dengan nilai rata-rata akurasi dan F1-Score hampir setara dengan model yang hanya menggunakan data historis, yakni sebesar 0,91.

7. Analisis *Threshold* Optimal

Threshold 1% menghasilkan terlalu banyak noise, yang terlihat dari variasi nilai akurasi dan F1-Score yang

lebih besar, dengan beberapa outlier yang menunjukkan performa model menjadi kurang konsisten. Oleh karena itu, *threshold* 1% tidak dianggap sebagai pilihan yang ideal. Di sisi lain, *threshold* 3% menunjukkan nilai rata-rata akurasi dan F1-Score yang hampir identik dengan *threshold* 2%. Namun, persebaran nilai pada *threshold* 3% sedikit lebih besar, sehingga stabilitasnya masih berada di bawah *threshold* 2%.

Meskipun performa *threshold* 3% tetap cukup stabil, hasil analisis secara keseluruhan menegaskan bahwa *threshold* 2% memberikan keseimbangan terbaik antara akurasi dan stabilitas. Oleh karena itu, *threshold* 2% dipilih sebagai nilai *threshold* optimal dalam penelitian ini

8. Analisis Korelasi Fitur Terkuat

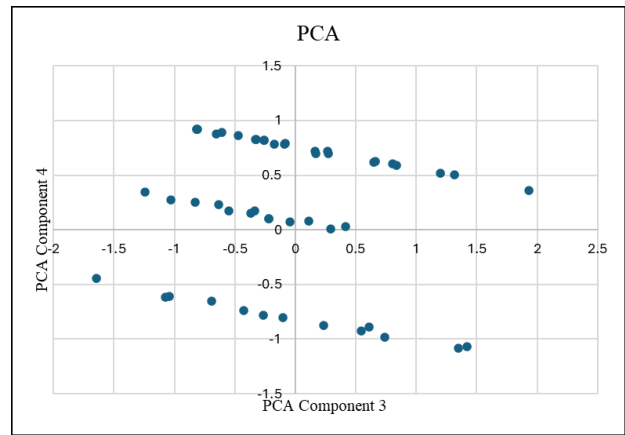
Berdasarkan analisis korelasi fitur untuk masing-masing sektor, ditemukan bahwa beberapa variabel memiliki korelasi yang lebih kuat dibandingkan variabel lainnya. Berikut adalah temuan utama dari analisis tersebut:

1. P/B Value variabel korelasi tertinggi secara konsisten di seluruh sektor, menunjukkan relevansinya dalam menggambarkan bagaimana pasar menilai perusahaan dibandingkan dengan nilai bukannya.
2. ROA memiliki korelasi signifikan di sektor barang baku dan energi. Sektor-sektor ini cenderung memiliki intensitas modal yang tinggi, sehingga profitabilitas yang dihasilkan dari aset perusahaan menjadi faktor yang lebih relevan dalam menilai kinerja perusahaan di sektor tersebut.
3. ROE memiliki korelasi kuat ditemukan di sektor bahan baku dan keuangan. Hal ini mengindikasikan bahwa profitabilitas terhadap ekuitas pemegang saham merupakan faktor penting untuk menilai kinerja perusahaan di kedua sektor ini.
4. DER memiliki korelasi negatif atau rendah, yang berarti rasio *leverage* ini kurang relevan untuk memprediksi pergerakan harga saham.

Berdasarkan hasil tersebut, variabel P/B Value, ROA, dan ROE dipilih sebagai fokus utama untuk eksperimen kombinasi fitur pada tahap selanjutnya. Meskipun korelasinya tidak terlalu besar, pendekatan ini tetap relevan mengingat model ANN yang akan digunakan dalam penelitian ini memiliki kemampuan menangkap pola hubungan non-linear.

9. Analisis Kombinasi Fitur

Gambar 4.9 menampilkan scatter plot hasil PCA, yang menggambarkan distribusi data setelah direduksi menjadi komponen utama.



GAMBAR 4.9
Scatter Plot Hasil PCA

Dari hasil PCA, diketahui bahwa empat komponen utama mampu menjelaskan 93.97% dari variansi data, dengan fitur-fitur seperti *Return*, *EPS*, *P/B Value*, dan *P/E Ratio* menjadi fitur yang paling signifikan pada setiap komponen utama. Penggunaan PCA memungkinkan model untuk menggabungkan informasi yang relevan ke dalam komponen yang lebih sedikit dan independen, serta mengurangi multikolinearitas antar fitur, yang berkontribusi terhadap peningkatan stabilitas model.

Untuk memastikan hasil analisis, dilakukan pengujian hipotesis untuk membandingkan performa model berbasis data historis dengan model yang menggunakan seluruh fitur dengan PCA. Pengujian ini bertujuan untuk memberikan bukti statistik yang mendukung kesimpulan bahwa data historis memiliki kontribusi yang lebih signifikan dibandingkan penambahan variabel fundamental.

10. Uji T

Uji t digunakan untuk menguji apakah rata-rata akurasi model berbasis data historis lebih besar dibandingkan model kombinasi fitur dengan PCA. Hipotesis yang digunakan adalah sebagai berikut:

- $H_0: \bar{x}_1 \leq \bar{x}_2$ (Rata-rata akurasi model berbasis data historis lebih kecil atau sama dengan rata-rata akurasi model kombinasi fitur dengan PCA)
- $H_1: \bar{x}_1 > \bar{x}_2$ (Rata-rata akurasi model berbasis data historis lebih besar dari rata-rata akurasi model kombinasi fitur dengan PCA)

Diperoleh Data:

- Ukuran sampel (n_1, n_2) : 19
- Rata-rata model pertama (\bar{x}_1) : 0.927
- Rata-rata model kedua (\bar{x}_2) : 0.905
- Varians model pertama (s_1^2) : 0.008
- Varians model kedua (s_2^2) : 0.011
- Tingkat signifikansi (α) : 0.05

Statistik Uji t:

$$t_{hitung} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\left(\frac{s_1^2}{n_1}\right) + \left(\frac{s_2^2}{n_2}\right)}} = \frac{0.927 - 0.905}{\sqrt{\left(\frac{0.008}{19}\right) + \left(\frac{0.011}{19}\right)}} = 0.695$$

Derajat Kebebasan (df):

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\left[\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1-1}\right] + \left[\frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2-1}\right]} = \frac{\left(\frac{0.008}{19} + \frac{0.011}{19}\right)^2}{\left(\frac{0.008}{19}\right)^2 + \left(\frac{0.011}{19}\right)^2} = 35.124$$

Nilai Kritis dan *p-value*:

Dengan tingkat signifikansi $\alpha = 0.05$ untuk uji satu sisi, nilai kritis dari tabel distribusi t adalah:

$$t_{tabel} = 1.69$$

p-value dalam analisis ini dihitung menggunakan Python dengan bantuan *library*, hasil perhitungan yaitu:

$$p - value = 0.2455$$

Daerah Kritis:

- Jika $t_{hitung} > t_{tabel}$, maka tolak H_0
- Jika $p - value < \alpha$, maka tolak H_0

Karena $t_{hitung} = 0.695 < t_{tabel} = 1.69$ dan $p - value > 0.05$, maka H_0 tidak dapat ditolak. Dengan demikian, tidak cukup bukti untuk menyatakan bahwa rata-rata akurasi model berbasis data historis lebih besar dari rata-rata akurasi model kombinasi fitur dengan PCA.

11. Uji F

Uji F digunakan untuk menguji apakah varians model historis lebih kecil dibandingkan PCA. Hipotesis yang digunakan adalah sebagai berikut:

- $H_0: s_1^2 \geq s_2^2$ (Varians model berbasis data historis lebih besar atau sama dengan varians model kombinasi fitur dengan PCA)
- $H_1: s_1^2 < s_2^2$ (Varians model berbasis data historis lebih kecil dari varians model kombinasi fitur dengan PCA)

Statistik Uji F:

$$F_{hitung} = \frac{s_2^2}{s_1^2} = 1.375$$

Derajat Kebebasan (*df*):

$$df_1 = n_1 - 1 = 18$$

$$df_2 = n_2 - 1 = 18$$

Nilai Kritis dan *p-value*:

Dengan tingkat signifikansi $\alpha = 0.05$ untuk uji satu sisi, nilai kritis dari tabel distribusi F adalah:

$$F_{tabel} = 2.217$$

p-value dalam analisis ini dihitung menggunakan Python dengan bantuan *library*, hasil perhitungan yaitu:

$$p - value = 0.253$$

Daerah Kritis:

- Jika $F_{hitung} > F_{tabel}$, maka tolak H_0
- Jika $p - value < \alpha$, maka tolak H_0

Karena $F_{hitung} = 1.375 < F_{tabel} = 2.217$ dan $p - value > 0.05$, maka H_0 tidak dapat ditolak. Dengan demikian tidak ada cukup bukti untuk menyatakan varians model berbasis data historis lebih kecil dari varians model kombinasi fitur dengan PCA.

Hasil uji statistik menunjukkan bahwa tidak ada cukup bukti untuk menyatakan bahwa model berbasis data historis memiliki rata-rata akurasi yang lebih besar dibandingkan model kombinasi fitur dengan PCA (berdasarkan uji t) maupun varians yang lebih kecil (berdasarkan uji F). Dengan demikian, meskipun model berbasis data historis

menunjukkan performa lebih tinggi dalam analisis awal, perbedaan tersebut tidak signifikan secara statistik, sehingga keduanya dianggap memiliki performa yang setara dalam hal akurasi dan stabilitas variansi.

V. KESIMPULAN

Penelitian ini berhasil menentukan *threshold* optimal untuk mengklasifikasikan pergerakan harga saham mingguan ke dalam tiga kategori: 1 (naik), 0 (stagnan), dan -1 (turun). Berdasarkan hasil eksperimen, *threshold* sebesar 2% dipilih sebagai yang paling optimal. Pilihan ini menghasilkan akurasi dan F1-Score yang lebih baik dibandingkan *threshold* lainnya, yaitu 1% dan 3%.

Model prediktif menggunakan metode Artificial Neural Network (ANN) menunjukkan bahwa data historis harga Hasil analisis *threshold* optimal menunjukkan bahwa pemilihan nilai *threshold* memiliki pengaruh signifikan terhadap kinerja model dalam memprediksi pergerakan harga saham. Dari Gambar 4.1 dan Gambar 4.2, terlihat bahwa *threshold* 2% memberikan rata-rata akurasi dan F1-Score terbaik dibandingkan *threshold* lainnya yaitu sekitar 0,9.

Threshold 1% menghasilkan terlalu banyak noise, yang terlihat dari variasi nilai akurasi dan F1-Score yang lebih besar, dengan beberapa outlier yang menunjukkan performa model menjadi kurang konsisten. Oleh karena itu, *threshold* 1% tidak dianggap sebagai pilihan yang ideal. Di sisi lain, *threshold* 3% menunjukkan nilai rata-rata akurasi dan F1-Score yang hampir identik dengan *threshold* 2%. Namun, persebaran nilai pada *threshold* 3% sedikit lebih besar, sehingga stabilitasnya masih berada di bawah *threshold* 2%.

Threshold 2% dipilih sebagai nilai *threshold* optimal dalam penelitian ini pergerakan harga saham jangka pendek. Model berbasis data historis mencapai rata-rata akurasi 0,92 dengan performa konsisten dan stabil. Sebaliknya, penambahan rasio keuangan tanpa penerapan PCA menghasilkan rata-rata akurasi lebih rendah, yaitu antara 0,5 hingga 0,7. Hal ini menunjukkan bahwa rasio keuangan memiliki korelasi yang rendah dengan pergerakan harga saham dalam jangka pendek dan tidak berdampak baik untuk memprediksi pergerakan harga saham pada periode waktu tersebut.

Penelitian ini juga menemukan bahwa penerapan Principal Component Analysis (PCA) pada kombinasi data historis dan rasio keuangan dapat meningkatkan akurasi hingga 0,91. Akurasi ini hampir setara dengan model berbasis data historis saja. Hasil uji statistik juga menunjukkan bahwa model berbasis data historis memiliki rata-rata akurasi yang lebih tinggi dibandingkan model kombinasi dengan PCA. Uji statistik juga menunjukkan bahwa varians keduanya tidak berbeda secara signifikan. Oleh karena itu, kedua model dapat dianggap memiliki performa yang setara dalam hal akurasi dan stabilitas variansi.

REFERENSI

- [1] M. Ali, D. M. Khan, M. Aamir, A. Ali, and Z. Ahmad, "Predicting the direction movement of financial time series using artificial neural network and support vector machine," *Complexity*, vol. 2021, pp. 1–13, 2021.

- [2] S. Patalay and M. R. Bandlamudi, "Stock price prediction and portfolio selection using artificial intelligence," *Asia Pacific Journal of Information Systems*, vol. 30, no. 1, pp. 31–52, 2020.
- [3] S. Schmidgall, R. Ziaei, J. Achterberg, L. Kirsch, S. P. Hajiseyedrazi, and J. K. Eshraghian, "Brain-inspired learning in artificial neural networks: A review," *APL Machine Learning*, vol. 2, no. 2, May 2024, doi: 10.1063/5.0186054.
- [4] A. Rose, "How do Artificial Neural Networks Work," *Journal of Advances in Science and Technology*, vol. 20, no. 1, pp. 172–177, 2024, doi: 10.29070/ttrkmm98.
- [5] P. Chhajer, M. Shah, and A. Kshirsagar, "The applications of artificial neural networks, support vector machines, and long short-term memory for stock market prediction," *Decision Analytics Journal*, vol. 2, p. 100015, 2022. DOI: 10.1016/j.dajour.2021.100015.
- [6] A. T. Oyewole, O. B. Adeoye, W. A. Addy, C. C. Okoye, O. C. Ofodile, and C. E. Ugochukwu, "Predicting stock market movements using neural networks: A review and application study," *Computer Science & IT Research Journal*, vol. 5, no. 3, pp. 651–670, 2024.
- [7] N. Ayyildiz and O. Iskenderoglu, "How effective is machine learning in stock market predictions?," *Heliyon*, vol. 10, no. 2, 2024.
- [8] A. R. Marakhimov and K. Khudaybergenov, "A Fuzzy MLP Approach for Nonlinear System Identification," *Journal of Mathematical Sciences*, vol. 265, no. 1, pp. 43–51, 2022, doi: 10.1007/s10958-022-06043-z.
- [9] X. Zhou, W. Zhang, Z. Chen, S. Diao, and T. Zhang, "Efficient Neural Network Training via Forward and Backward Propagation Sparsification," *arXiv: Learning*, 2021.
- [10] V. Andreieva and N. Shvai, "Generalization of cross-entropy loss function for image classification," *Journal of Computer Vision Applications*, vol. 3, pp. 3–10, 2021, doi: 10.18523/2617-7080320203-10.
- [11] J. Elis, M. P. B. Kellen, and G. S. Tefa, "Pengaruh Kinerja Keuangan Terhadap *Return* Saham Pada Perusahaan Manufaktur Sektor Aneka Industri Yang Terdaftar di Bursa Efek Indonesia Periode Tahun 2018–2022," *Menawan*, vol. 2, no. 5, pp. 122–142, 2024, doi: 10.61132/menawan.v2i5.809.
- [12] J. Bocianowski, D. Wrońska-Pilarek, A. Kryzstofiak-Kaniewska, K. Matusiak, and B. Wiatrowska, "Comparison of Pearson's and Spearman's correlation coefficients values for selected traits of *Pinus sylvestris* L.," *Research Square*, 2024, doi: 10.21203/rs.3.rs-4380975/v1.
- [13] C. Thirumalai, S. A. Chandhini, and M. Vaishnavi, "Analysing the concrete compressive strength using Pearson and Spearman," in *Proc. 2017 Int. Conf. Electronics, Communication and Aerospace Technology (ICECA)*, Coimbatore, India, 2017, pp. 215–218, doi: 10.1109/ICECA.2017.8212799.
- [14] A. J. Culyer, "Spearman's Rank Correlation Coefficient," *Handbook of Statistical Methods*, 2014.
- [15] D. Qiao, X. Chen, and B. Huang, "Principal Component Analysis," in *Introduction to Data Science Techniques*, 263–273, 2024, doi: 10.1016/b978-0-443-15928-2.00007-0.
- [16] I. Świetlicka, W. Kuniszyk-Józkowiak, and M. Świetlicki, "Artificial Neural Networks Combined with the Principal Component Analysis for Non-Fluent Speech Recognition," *Sensors*, vol. 22, no. 1, pp. 321–341, 2022, doi: 10.3390/s22010321.
- [17] F. J. Gravetter and L. B. Wallnau, *Statistics for the Behavioral Sciences*, 10th Edition. 2016.
- [18] R. E. Walpole, R. H. Myers, S. L. Myers, and K. Ye, *Probability and Statistics for Engineers and Scientists* (9th Edition). 2011.
- [19] J. Nugroho, L. Pangaribuan, and W. Prastio, "Pengaruh ROA, ROE, dan EPS terhadap harga saham perusahaan perbankan yang terdaftar pada BEI periode 2018–2022," *Jurnal Akuntansi*, vol. 13, no. 2, 2024, doi: 10.46806/ja.v13i2.1167.
- [20] J. Irawati and M. Andriani, "Pengaruh DER, ROA dan EPS terhadap PBV," *Business Management Review*, vol. 2, no. 1, pp. 241–440, 2012.
- [21] A. M. F. Sajeetha, M. F. Nusaika, and M. N. F. N. Safana, "An Empirical Study on Determinants of Price Earnings Ratio: Evidence from Listed Food, Beverage and Tobacco Companies in Colombo Stock Exchange," *Asian Journal of Economics, Business and Accounting*, vol. 23, no. 10, pp. 32–43, 2023, doi: 10.9734/ajeba/2023/v23i10968.
- [22] S. Sufriani and M. Rimawan, "Analisis *Return* on Equity dan Debt to Equity Ratio," *Owner*, vol. 4, no. 2, pp. 308–316, 2020, doi: 10.33395/owner.v4i2.228.