

Perbandingan Algoritma Support Vector Machine dan Decision Tree untuk Klasifikasi Performa Perusahaan

Mario Utomo, Rastri Prathivi*

Fakultas Teknologi Informasi dan Komunikasi, Teknik Informatika, Universitas Semarang, Semarang, Indonesia

Email: ¹marr.utm2002@gmail.com, ^{2,*}vivi@usm.ac.id

Email Penulis Korespondensi: vivi@usm.ac.id

Submitted: 04/06/2024; Accepted: 23/06/2024; Published: 23/06/2024

Abstrak—Dengan meningkatnya jumlah investor bursa saham pada akhir Desember 2023 mencapai 5,34 juta yang didominasi oleh investor generasi milenial menunjukkan bahwa pelaku pasar modal benar-benar percaya terhadap fundamental dan prospek ekonomi di Indonesia. Kurangnya literasi keuangan pada generasi milenial ini seringkali menyebabkan investasi menjadi kurang efektif dan berisiko tinggi. Banyak generasi milenial yang memilih suatu saham berdasarkan tren jangka pendek atau rekomendasi yang kurang analisis. Dalam mengatasi masalah ini, diperlukan analisis yang lebih terstruktur dalam pemilihan suatu saham. Salah satu metode yang dapat digunakan adalah klasifikasi performa suatu perusahaan berdasarkan kinerjanya menggunakan berbagai indikator dan rasio keuangan. Dikarenakan kinerja performa suatu perusahaan menjadi faktor yang mempengaruhi pergerakan nilai sahamnya. Penelitian ini akan membandingkan *Support Vector Machine* dan *Decision Tree* dengan pendekatan *One Against All* dalam mengklasifikasikan performa perusahaan. *Feature* yang digunakan untuk klasifikasi performa perusahaan ini terdiri dari 3 rasio keuangan yaitu rasio profitabilitas (ROA), likuiditas (CR), dan leverage (DER). *Labelling* atau target dalam klasifikasi dibagi menjadi 3 kategori yaitu normal, baik, dan kurang baik. Pada penelitian ini akan mempertimbangkan evaluasi seperti *accuracy*, *cross validation*, dan *confusion matrix*. Hasil kinerja algoritma *Support Vector Machine* menghasilkan akurasi sebesar 86,67%, sedangkan pada algoritma *Decision Tree* menghasilkan akurasi sebesar 93,33%. Sehingga algoritma *Decision Tree* menghasilkan hasil yang lebih akurat dibandingkan algoritma *Support Vector Machine* dalam klasifikasi.

Kata Kunci: Klasifikasi; Decision Tree; Support Vector Machine; Keuangan; Performa

Abstract—The number of stock exchange investors in Indonesia reached 5.34 million by the end of December 2023. This figure is dominated by millennial generation investors, indicating a growing confidence in the fundamentals and economic prospects of the Indonesian capital market. However, the lack of financial literacy among this generation often results in ineffective and high-risk investments. Many millennials choose stocks based on short-term trends or recommendations that lack analysis. To address this issue, a more structured approach to stock selection is required. One method that can be employed is the classification of a company's performance based on its performance using various financial indicators and ratios. As the performance of a company affects the movement of its stock value, this research will compare Support Vector Machine and Decision Tree with the One Against All approach in classifying company performance. The features used for the classification of company performance consist of three financial ratios: profitability (ROA), liquidity (CR), and leverage (DER). The labels or targets in the classification are divided into three categories: normal, good, and unfavorable. This research will consider evaluations such as accuracy, cross validation, and confusion matrix. The results of the Support Vector Machine (SVM) algorithm demonstrated an accuracy of 86.67%, while the Decision Tree (DT) algorithm exhibited an accuracy of 93.33%. Consequently, the DT algorithm produced more accurate results than the SVM algorithm in classification.

Keywords: Classification; Decision Tree; Support Vector Machine; Finance; Performance

1. PENDAHULUAN

Jumlah investor di pasar modal Indonesia mencapai 12,16 juta orang. Jika dibandingkan dengan Desember 2022, jumlah investor di pasar modal meningkat sebesar 18,01% pada Desember 2023. Dari jumlah total investor pasar modal di Indonesia di antaranya terdapat sekitar 5,34 juta investor menginvestasikan dananya di pasar saham. Meningkatnya jumlah investor di pasar modal menunjukkan bahwa pelaku pasar modal mempunyai keyakinan penuh terhadap prinsip fundamental dan potensi masa depan perekonomian Indonesia [1]. Bursa Efek Indonesia (BEI) merupakan pasar modal di Indonesia yang memfasilitasi interaksi investor dan perusahaan melalui pertukaran instrumen keuangan, termasuk perdagangan pada bursa saham [2]. Investor generasi milenial sangat menyukai saham, salah satu instrumen investasi yang paling populer di pasar modal karena tingkat pengembaliannya yang tinggi [3]. Dalam berinvestasi investor didorong untuk berpartisipasi karena tingkat keuntungannya, namun mereka cenderung menghindari mengambil terlalu banyak resiko [4]. Kurangnya literasi keuangan pada generasi milenial ini seringkali menyebabkan investasi menjadi kurang efektif dan berlebihan. Banyak generasi milenial yang memilih suatu saham berdasarkan tren jangka pendek atau rekomendasi yang kurang analisis. Kondisi tersebut menunjukkan perlunya pendekatan pemilihan saham yang lebih terstruktur, hal ini akan membantu investor muda dalam mengambil keputusan yang lebih tepat berdasarkan data yang akurat. Investor dapat menggunakan hasil analisis sebagai ukuran kinerja performa perusahaan dalam periode waktu tertentu [5]. Dengan metode analisis fundamental dapat digunakan untuk memprediksi kinerja suatu perusahaan.

Analisis fundamental adalah analisis data siklus bisnis dan laporan keuangan yang terkait dengan situasi internal perusahaan, termasuk faktor ekonomi, keuangan, dan lain-lain yang berdampak pada perusahaan [6]. Laporan keuangan memberikan informasi terkait kinerja dan kondisi keuangan pada perusahaan. Kinerja keuangan pada laporan keuangan perusahaan merujuk pada penilaian kesehatan keuangan suatu perusahaan mulai dari harta, utang, modal serta hasil yang dihasilkan dan efisiensi operasional perusahaan [7]. Rasio keuangan atau *indeks* keuangan

adalah tolak ukur yang digunakan untuk mempermudah para investor dalam menilai suatu kinerja keuangan perusahaan [8]. Prospek dan kemampuan perusahaan dalam menjalankan suatu bisnis di masa depan dapat dicerminkan melalui performa keuangan perusahaan itu sendiri [9]. Dalam hal ini bagi seorang investor saham harus lebih aktif dalam melihat perkembangan kinerja performa perusahaan, yaitu dalam hal analisis rasio keuangan guna memberikan informasi terkait performa perusahaan. Memahami kondisi perusahaan tersebut membantu para investor dalam mengetahui performa perusahaan dalam kondisi yang normal, baik, atau kurang baik.

Istilah *data mining* digunakan untuk menguraikan atau membagi penemuan pengetahuan yang ditemukan dalam suatu kumpulan data [10]. Metode *machine learning* cocok dalam melakukan pemantauan, pengelolaan, kebijakan berkelanjutan, dan pembuatan [11]. Dalam pembelajaran *machine learning*, klasifikasi merupakan salah satu metode untuk melakukan *data mining* dan membantu dalam pengelompokan berdasarkan karakteristik khusus [12]. Klasifikasi performa perusahaan dilakukan untuk berbagai alasan yang penting bagi pemangku kepentingan yang berbeda, seperti manajemen, investor, kreditur, regulator, dan analisis keuangan. Klasifikasi ini akan membagi perusahaan dalam berbagai kondisi kinerja keuangan normal, baik, dan kurang baik. Dalam penelitian ini, algoritma pembelajaran mesin seperti *Support Vector Machine* dan *Decision Tree* akan dibandingkan untuk melakukan klasifikasi kinerja performa suatu perusahaan.

Penelitian lain telah membandingkan algoritma pembelajaran mesin *Support Vector Machine* dan *Decision Tree* dalam klasifikasi kumpulan data yang kompleks [13]. Untuk membedakan penyakit kanker paru-paru, penelitian ini menggabungkan metode *Support Vector Machine* dan *Decision Tree*. Selama pelaksanaan perbandingan ini, peneliti melakukan berbagai eksperimen dengan masing-masing algoritma dan mencatat keakuratan hasilnya. Hasil studi ini menghasilkan algoritma *Support Vector Machine* mempunyai tingkat akurasi yang tinggi yaitu sebesar 62,3%, dan pemanfaatan *forward Selection* meningkatkan akurasi algoritma tersebut. Selain itu penelitian serupa juga pernah dilakukan [14], penelitian ini juga membandingkan algoritma *Support Vector Machine* dan *Decision Tree*. Penelitian ini bertujuan untuk menilai efektivitas metode *Support Vector Machine* dan *Decision Tree* dalam klasifikasi bayi yang berpotensi stunting. Perbandingan ini dimaksudkan untuk memberikan hasil yang tepat dan efektif, serta merekomendasikan cara untuk mengenali bayi yang berisiko stunting. Hasil studi ini menghasilkan model *Support Vector Machine* memiliki akurasi yang lebih tinggi yaitu sebesar 83%, sedangkan pada model *Decision Tree* memiliki akurasi yang lebih rendah yaitu sekitar 78%. Kedua model tersebut memiliki *confusion matrix* yang tinggi dalam mengenali balita yang berisiko stunting. Dalam penelitian ini [15] juga membandingkan algoritma pembelajaran mesin *Support Vector Machine* dan *Decision Tree* dengan menggunakan data penyakit diabetes. Penelitian ini menunjukkan bahwa akurasi pada algoritma *Decision Tree* mencapai 85,93%, sedangkan akurasi pada algoritma *Support Vector Machine* sekitar 83,85%.

Selain penelitian yang berkonsentrasi pada perbandingan algoritma *data mining*, banyak penelitian juga membahas metode *cross-validation*, seperti yang dilakukan [16]. Penelitian tersebut menggunakan algoritma pembelajaran mesin *Support Vector Machine* dengan menggunakan data pasien terindikasi penyakit malaria. Tujuan penelitian ini adalah untuk menemukan dan mengklasifikasikan penyakit malaria berat menggunakan data historis pemeriksaan pasien. Hasil studi ini diperoleh kinerja algoritma *Support Vector Machine* menghasilkan akurasi 69,4% sebelum dilakukan normalisasi dan *cross validation*, setelah dilakukan normalisasi dan *cross validation* menghasilkan akurasi 92,3%. Hal ini menunjukkan bahwa *cross validation* memiliki pengaruh penting dalam meningkatkan hasil akhir akurasi.

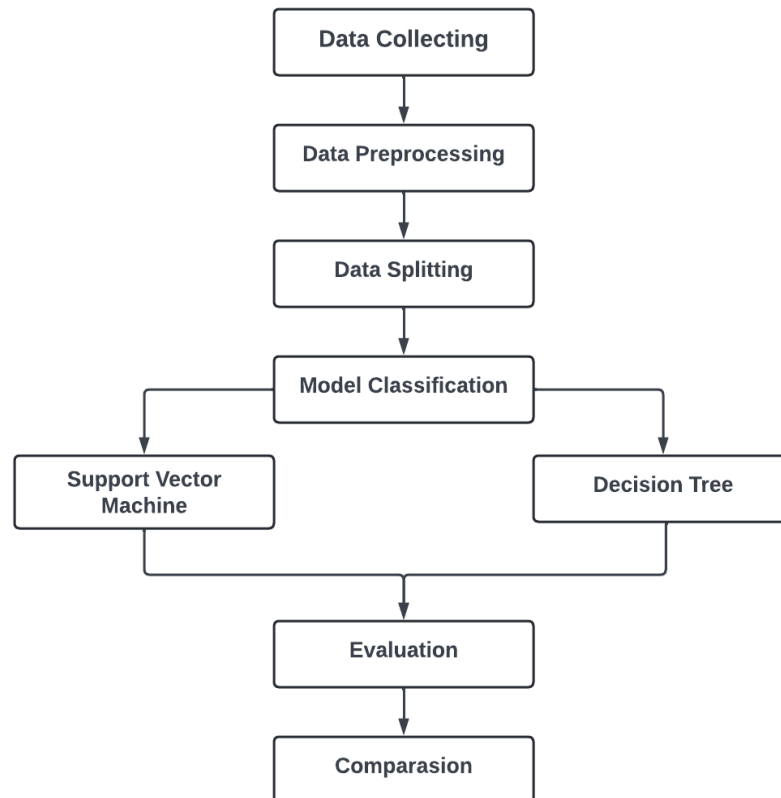
Selanjutnya penelitian yang dilakukan [17] yaitu melakukan implementasi algoritma SVM untuk klasifikasi *multiclass* pada data pasien yang menderita penyakit tiroid. Tujuannya adalah untuk klasifikasi *multiclass* dengan kategori pasien dengan hipertiroidisme, hipotiroidisme, dan normal. Pada penelitian ini menggunakan pendekatan *One Against One* dan *One Against All*. Penelitian ini menemukan bahwa pendekatan *One Against One* memiliki hasil yang lebih baik daripada pendekatan *One Against All*. Hal ini dikarenakan pendekatan *One Against One* memiliki nilai akurasi sebesar 99,53% dibandingkan dengan algoritma *One Against All*.

Sehubungan dengan permasalahan diatas, maka pada penelitian ini akan melakukan penelitian yang membandingkan akurasi dari algoritma *Decision Tree* dan *Support Vector Machine* dalam mengklasifikasikan performa perusahaan dengan pendekatan metode *multiclass One Against All*. Dataset yang digunakan pada laporan keuangan perusahaan, terdiri dari 150 data perusahaan, tiga *feature* karakteristik, dan tiga *labelling* atau target kelas. Metode pengujian dan evaluasi menggunakan *confusion matrix*. Diharapkan penelitian ini akan memberikan informasi tentang performa terbaik, yang akan membantu menghasilkan klasifikasi yang lebih akurat untuk mengklasifikasikan kinerja performa suatu perusahaan.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Melalui penelitian tentang perbandingan algoritma pembelajaran mesin *Support Vector Machine* dan *Decision Tree* dalam mengklasifikasikan performa perusahaan. Pada Gambar 1 akan menunjukkan langkah-langkah penting dan menarik yang akan dilakukan, yang akan membantu menghasilkan klasifikasi yang lebih akurat untuk mengklasifikasikan kinerja performa perusahaan.



Gambar 1. Alur Tahapan Penelitian

Pada Gambar 1, penelitian ini dimulai dengan *data collecting* atau pengambilan dataset dari laporan keuangan perusahaan pada Bursa Efek Indonesia. Setelah mendapatkan dataset untuk laporan keuangan perusahaan, langkah berikutnya adalah memproses data sebelum digunakan untuk menghindari duplikat, kehilangan nilai, dan standardisasi. Kemudian pemilihan *feature* yang akan digunakan dan menentukan *labelling* (target) untuk klasifikasi. Data yang telah melewati tahap *data understanding* akan memiliki kualitas dan siap untuk dilakukan *modelling* menggunakan algoritma pembelajaran mesin. Sebelum memasuki tahap *model classification* data yang telah melalui tahap *data understanding* akan dilakukan proses *splitting data* atau pemisahan data menjadi data *testing* (pengujian) dan data *training* (pelatihan) secara *random*.

Langkah berikutnya adalah pembuatan model menggunakan algoritma pembelajaran mesin. Algoritma *Support Vector Machine* dan *Decision Tree* akan digunakan dalam mengkategorikan performa perusahaan berdasarkan *feature* dan *labelling* (target) yang telah ditentukan. Performa perusahaan akan dikategorikan menjadi 3 kategori yaitu normal, baik, dan kurang baik. Kemudian penelitian dilanjutkan dengan tahap *evaluation* dengan *confusion matrix* seperti *accuracy*, *recall*, *precision*, dan *F1-score*. Pada tahap validasi akan digunakan *cross validation*. Langkah terakhir adalah dengan melakukan *comparasion* akurasi dan hasil evaluasi antara kedua algoritma pembelajaran mesin tersebut.

a. Data Collecting

Pada penelitian ini, dataset diambil dari *website* Bursa Efek Indonesia (BEI). Dataset ini berisikan laporan keuangan tahunan dari berbagai perusahaan yang terdaftar di pasar modal Indonesia. Dalam dataset ini, terdapat data sebanyak 150 baris dan 20 kolom.

b. Data Preprocessing

Data preprocessing merupakan metode untuk menggali informasi dari data sebelum data memasuki tahap *modelling*. Adapun langkah-langkah yang akan dilakukan selama *data preprocessing* sebagai berikut:

1. *Data Understanding* : Proses untuk menemukan, mengoreksi, dan menghapus kesalahan dalam suatu kumpulan data. Tujuannya untuk memastikan bahwa data yang akan digunakan dalam proses *modelling data*, analisis, atau pemrosesan lebih lanjut menjadi lebih akurat, lengkap, konsisten, dan relevan.
2. *Features and Labels Selection* : Sebelum tahap *model classification* menggunakan algoritma pembelajaran mesin, akan terlebih dahulu dilakukan proses pemilihan *feature* dan *labelling* (target).

c. Data Splitting

Data yang digunakan dalam membuat model dibagi secara *random* menggunakan *train test split* menjadi data *training* (pelatihan) dan data *testing* (pengujian) dengan perbandingan 80:20. Selanjutnya data dilakukan standarisasi menggunakan *StandardScaler*.

d. Model Classification

Model classification menggambarkan proses dimana sistem mengkategorikan atau mengelompokkan sebuah data ke dalam kelas yang berbeda berdasarkan karakteristiknya [18]. Dua pendekatan yang paling umum digunakan untuk klasifikasi dan prediksi adalah algoritma pembelajaran mesin *Support Vector Machine* dan algoritma *Decision Tree*. Dalam algoritma pembelajaran mesin *Support Vector Machine* dan *Decision Tree* memiliki kelebihan dalam generalisasi data yang tinggi dengan model klasifikasi yang lebih baik dengan menggunakan jumlah data yang relatif sedikit [19].

e. Support Vector Machine

Dalam klasifikasi dan regresi, *Support Vector Machine* adalah algoritma pembelajaran mesin yang digunakan untuk menemukan *hyperplane* yang memisahkan kelas-kelas data dalam ruang fitur dengan margin maksimum. Dalam masalah klasifikasi *biner* atau *multiclass*, Algoritma pembelajaran mesin seperti *Support Vector Machine* sangat efektif [20]. Pada tahapan pengujian *Support Vector Machine* ini akan dilakukan pendekatan *One Against All* dengan persamaan sebagai berikut:

1. Fungsi Kernel RBF

Di mana γ adalah parameter yang menentukan jangkauan pengaruh dari satu contoh pelatihan.

$$RBF = K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2) \quad (1)$$

2. Fungsi One Against All SVM

Untuk setiap kelas k (dari 1 hingga K), latih binary SVM classifier $f_k(x)$ untuk membedakan kelas k dari semua kelas lainnya. Di mana $\phi(x)$ adalah fungsi yang memetakan fitur asli ke ruang fitur yang lebih tinggi dengan kernel RBF.

$$One\ Against\ All = \frac{\min}{w_k b_k} \frac{1}{2} \|w_k\|^2 + C \sum_{i=1}^n \max(0, 1 - y_i^{(k)} (w_k \cdot \phi(x_i) + b_k)) \quad (2)$$

f. Decision Tree

Karena hasilnya yang dapat dipahami dalam bentuk kaidah keputusan, *Decision Tree* untuk klasifikasi adalah salah satu algoritma klasifikasi yang paling banyak digunakan dalam kedua klasifikasi dan regresi. Pohon keputusan mempartisi data berulang kali untuk memaksimalkan pemisahan, sampai setiap partisi hanya mengandung satu kelas atau tidak dapat dibagi lebih lanjut. Pada tahapan pengujian *Decision Tree* ini akan dilakukan dengan pendekatan *One Against All* dengan persamaan sebagai berikut:

1. Fungsi Entropy

Entropi (H) digunakan untuk mengukur ketidakpastian atau impuritas suatu himpunan data. Di mana p_i adalah probabilitas kemunculan kelas i dalam himpunan data D , dan k adalah jumlah kelas.

$$H(D) = - \sum_{i=1}^k p_i \log_2(p_i) \quad (3)$$

2. Gain

Informasi gain (IG) mengukur penurunan entropi dari suatu himpunan data setelah dibagi berdasarkan suatu atribut. Di mana D_v adalah himpunan data dengan nilai atribut A sama dengan v

$$IG(D, A) = H(D) - \sum_{v \in \text{Values}(A)} \frac{|D_v|}{|D|} H(D_v) \quad (4)$$

3. Fungsi One Against All

Setiap kelas k (dari 1 hingga K), latih binary classifier $f_k(x)$ untuk membedakan kelas k untuk membedakan kelas k dari semua kelas lainnya. Misalkan y adalah prediksi.

$$\hat{y} = \arg \max_k f_k(x) \quad (5)$$

g. Evaluation

Proses mengukur hasil evaluasi dari model yang telah di implementasikan pada tahap *modelling* dikenal sebagai evaluasi [21]. Studi ini melihat model klasifikasi yang menggunakan *confusion matrix*. Matriks yang digunakan dalam penelitian ini mencakup *recall*, *precision*, dan *F1-Score*. Nilai *accuracy*, *precision*, dan *recall* dapat dihitung berdasarkan matriks persamaan:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (6)$$

$$Precision = \frac{TP}{TP+FP} \quad (7)$$

$$Recall = \frac{TP}{TP+FN} \quad (8)$$

$$f1 - score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (9)$$

$$F1_{macro} = \frac{1}{K} \sum_{k=1}^K F1_k \quad (10)$$

h. Comparasion

Dalam konteks klasifikasi, *comparasion* biasanya merujuk pada proses membandingkan performa berbagai model klasifikasi atau algoritma untuk menentukan model mana yang paling efektif dalam memprediksi kelas target. *Comparasion* dapat melibatkan berbagai matrik evaluasi, teknik validasi, dan visualisasi hasil untuk memastikan bahwa pemilihan model didasarkan pada analisis yang komprehensif dan menyeluruh. Tujuan dari *comparasion* adalah untuk memilih model terbaik yang memiliki akurasi tinggi, generalisasi yang baik, dan performa yang konsisten pada data baru.

3. HASIL DAN PEMBAHASAN

3.1 Dataset

Penelitian ini menggunakan dataset laporan keuangan tahunan perusahaan yang terdaftar di pasar modal Indonesia dan dapat diakses pada *website* Bursa Efek Indonesia. Dataset berisikan 150 baris dan 20 kolom, baris berisikan data 150 perusahaan dari berbagai sektor dan kolom berisikan atribut dari indikator keuangan. Dataset ini mengandung dua puluh atribut atau fitur penting yang digunakan dalam analisis, yaitu:

- a. Kode perusahaan yang *go public* pada bursa saham disebut *stocks*.
- b. Lingkungan suatu usaha bisnis disebut sektor.
- c. *Total Assets* adalah jumlah aset yang dimiliki perusahaan.
- d. *Total Current Assets* adalah jumlah aset lancar yang dimiliki perusahaan.
- e. *Total Equity* adalah jumlah hak pemilik dalam perusahaan setelah dikurangi dengan jumlah semua kewajiban perusahaan.
- f. *Total Liabilities* adalah jumlah hutang dan semua kewajiban yang dimiliki perusahaan.
- g. *Total Current Liabilities* adalah jumlah kewajiban jangka pendek perusahaan yang berlangsung kurang dari dua belas bulan.
- h. *Revenue* adalah jumlah pendapatan kotor yang dihasilkan perusahaan dari kegiatan operasional, penjualan, dan aktivitas perusahaan.
- i. *Net Income* adalah jumlah pendapatan perusahaan setelah dikurangi dengan semua biaya dan pengeluaran dari pendapatan kotor.
- j. *Current Share* adalah jumlah lembar saham yang beredar.
- k. *Book Value* adalah nilai sebuah aset dengan sejumlah penyusutan nilai yang dibebankan selama umur penggunaan aset.
- l. *Last Price* adalah harga terakhir saham yang berada di bursa saham.
- m. PBV adalah rasio harga saham terhadap nilai bukunya.
- n. EPS adalah rasio yang menghitung pendapatan bersih suatu perusahaan selama satu tahun dengan membagi jumlah lembar saham yang beredar rata-rata.
- o. PER adalah rasio harga saham dengan laba per lembar saham.
- p. DER adalah rasio jumlah hutang terhadap jumlah ekuitas perusahaan.
- q. ROE adalah rasio jumlah pendapatan bersih dengan jumlah ekuitas perusahaan.
- r. ROA adalah rasio jumlah pendapatan bersih dengan jumlah aset perusahaan.
- s. CR atau *Current Ratio* adalah rasio keuangan yang menghitung seberapa baik perusahaan dapat membayar hutang jangka pendeknya dengan aset lancarnya..
- t. NPM adalah perbandingan total laba bersih perusahaan dengan total pendapatan perusahaan.

3.2 Data Preprocessing

Dalam proses Data Preprocessing ini, memahami dataset laporan keuangan perusahaan untuk klasifikasi performa perusahaan. Cara untuk menganalisis dataset sebagai berikut:

- a. Data Understanding

Tabel 1. Informasi Dataset

Column	Non-Null Count	Dtype
STOCKS	150 non-null	object
SEKTOR	150 non-null	object
TOTAL ASSETS	150 non-null	int64
TOTAL CURRENT ASSETS	150 non-null	int64
TOTAL EQUITY	150 non-null	int64
TOTAL LIABILITIES	150 non-null	int64
TOTAL CURRENT LIABILITIES	150 non-null	int64
REVENUE	150 non-null	int64
NET INCOME	150 non-null	int64
CURRENT SHARE	150 non-null	int64

BOOK VALUE	150 non-null	int64
LAST PRICE	150 non-null	int64
PBV	150 non-null	float64
EPS	150 non-null	float64
PER	150 non-null	float64
DER	150 non-null	float64
ROE	150 non-null	float64
ROA	150 non-null	float64
CR	150 non-null	float64
NPM	150 non-null	float64

Pada Tabel 1 menunjukkan 20 atribut dan 150 baris data. Dari semua atribut sudah sesuai dengan tipe data tersebut. Terdapat 2 atribut kategorikal dan 18 atribut numerik. Atribut kategorikal yaitu stocks dan sektor. Atribut numerik yaitu total assets, total current assets, total equity, total liabilities, total current liabilities, revenue, net income, current share, book value, last price, PBV, EPS, PER, DER, ROE, ROA, CR, dan NPM. Setelah dilakukannya *preprocessing* data dataset diatas tidak ditemukan *missing value* dan *duplicate data*.

b. *Feature dan Labelling*

Pada tahap ini akan menentukan *feature* dan *labelling* (target) sebelum dilanjutkan ke tahap modelling klasifikasi menggunakan *Support Vector Machine* dan *Decision Tree*. *Feature* yang akan digunakan dalam penelitian ini adalah atribut DER, ROA, dan CR pada laporan keuangan. Pemilihan *feature* tersebut dikarenakan rasio profitabilitas (ROA), likuiditas (CR), dan leverage (DER) berpengaruh terhadap nilai performa sebuah perusahaan [22]. *Labelling* atau target pada penelitian ini dibagi menjadi 3 kategori klasifikasi yaitu dengan perusahaan dengan performa normal, baik, dan kurang baik. Performa perusahaan normal dengan nilai DER 100% – 200%, ROA dengan nilai 5% – 10%, dan CR dengan nilai 1,5 kali – 2 kali. Performa perusahaan baik yaitu dengan nilai DER kurang dari 100%, ROA lebih dari 10%, dan CR lebih dari 2 kali. Performa perusahaan kurang baik yaitu dengan nilai DER lebih dari 200%, ROA kurang dari 5%, dan CR kurang dari 1,5 kali.

3.3 Data Splitting

```

[62] X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

[63] X_train.shape , X_test.shape
      ((120, 3), (30, 3))

[64] y_train.shape , y_test.shape
      ((120,), (30,))
    
```

Gambar 2. Implementasi Data Splitting

Pada Gambar 2 menunjukkan tahapan dalam melakukan proses *splitting data*. *Splitting data* dilakukan secara *random* dengan menggunakan *library train_test_split* dengan perbandingan 80:20. *Splitting data* tersebut menghasilkan jumlah *x_train* 120 sampel dengan 3 fitur, *x_test* menghasilkan 30 sampel dan 3 fitur. Data *y_train* berisikan 120 sampel dan data *y_test* berisikan 30 sampel. Memiliki lebih banyak data *training* dibandingkan dengan data *testing* membantu dalam membangun model yang lebih kuat dan dapat diandalkan. Dalam membantu dalam mengoptimalkan performa algoritma, meningkatkan stabilitas numerik, membantu dalam deteksi anomali, dan mempermudah interpretasi model dilakukanlah proses standarisasi pada data *x_train* dan *x_test* menggunakan *library StandardScaler*.

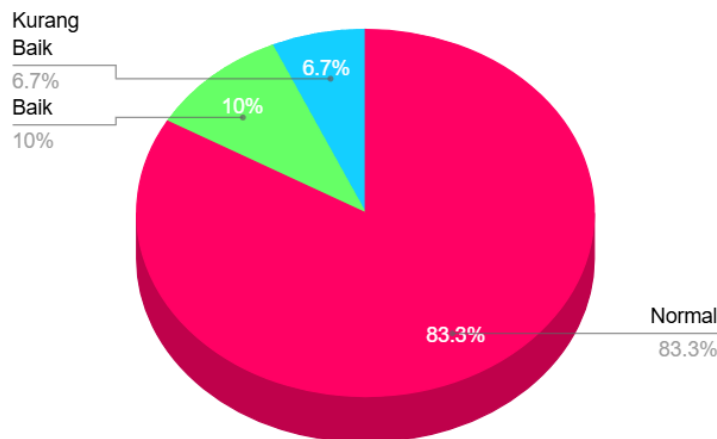
3.4 Pengujian Support Vector Machine

Pada tahap pengujian ini, dataset yang telah melalui langkah-langkah *preprocessing*, pemilihan *feature labelling*, dan *splitting data* akan diselidiki lebih lanjut menggunakan teknik pengujian menggunakan algoritma pembelajaran mesin *Support Vector Machine*. Pada algoritma *Support Vector Machine* ini akan melakukan pendekatan *multiclass One Against All* atau *One Against Rest*. Nilai *accuracy*, *recall*, *F1-score*, *precision*, dan *confusion matrix* akan dibandingkan selama tahap pengujian. Untuk melakukan validasi klasifikasi akan menggunakan metode *cross validation*. Berikut adalah hasil yang didapatkan dengan pengujian dengan komposisi data *training* (pelatihan) dan data *testing* (pengujian) dengan perbandingan data 80:20.

Tabel 1. Hasil Distribusi SVM *One Against All*

Normal	Baik	Kurang Baik
25	3	2

Pie Chart

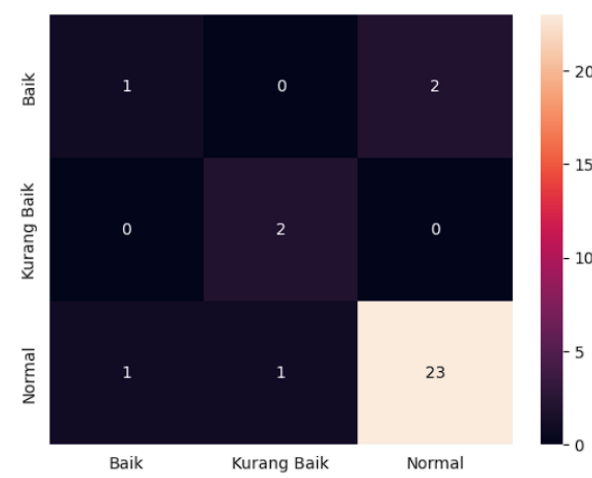


Gambar 3. Distribusi Klasifikasi SVM

Berdasarkan hasil Tabel 2 dan Gambar Grafik 3 diatas digunakan 30 data pelatihan yang dapat disimpulkan bahwa 25 perusahaan termasuk dalam kategori normal atau sebanyak 83%. Dan 3 perusahaan atau 7 persen termasuk kategori baik. Sisanya sebanyak 3 perusahaan atau 10% termasuk dalam kategori kurang baik. Hasil dari algoritma *Support Vector Machine* mengenai *accuracy*, *F1-score*, *precision*, *recall*, *K-Fold Validation* dan *confusion matrix* berikut ini.

Tabel 3. Hasil Kinerja SVM

Accuracy	Precision	Recall	F1-score	K-Fold Validation
86,67%	69,56%	75,11%	70,67%	90%



Gambar 4. Confusion Matrix SVM

Berdasarkan hasil Tabel 3 dan Gambar 4 diatas setelah dilakukannya pengujian menggunakan algoritma *Support Vector Machine* dan menghasilkan nilai yang meliputi *accuracy*, *precision*, *F1-score*, *recall*, *K-fold validation* dan *confusion matrix* dari *Support Vector Machine*. Maka hasil dari pengujian tersebut memiliki tingkat nilai *accuracy* sebesar 86,67%, nilai *precision* sebesar 69,56%, nilai *recall* sebesar 75,11%, nilai *F1-score* sebesar 70,67%, dan hasil *accuracy* setelah dilakukannya *cross-validation* dengan k=5 sebesar 90%.

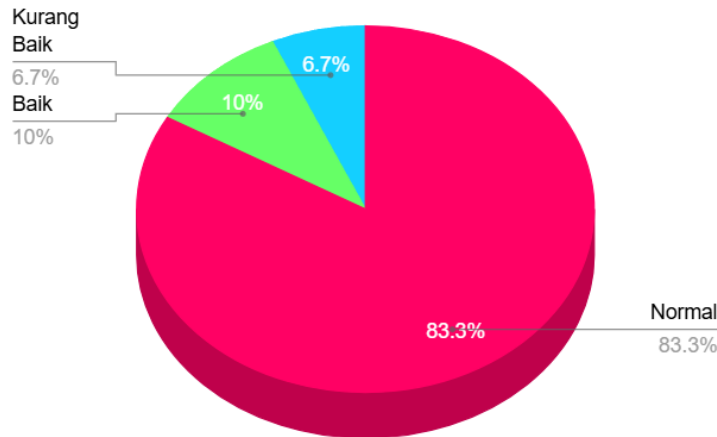
3.5 Pengujian Decision Tree

Pada tahap pengujian ini, dataset yang sudah melalui langkah-langkah *preprocessing*, pemilihan *feature labelling*, dan *splitting data* akan diselidiki lebih lanjut menggunakan teknik pengujian menggunakan algoritma pembelajaran mesin *Decision Tree*. Pada algoritma *Decision Tree* ini akan melakukan pendekatan *multiclass One Against All* atau *One Against Rest*. Nilai *accuracy*, *recall*, *F1-score*, *precision*, dan *confusion matrix* akan dibandingkan selama tahap pengujian. Untuk validasi klasifikasi akan menggunakan metode *cross validation*. Berikut adalah hasil yang didapatkan dengan pengujian dengan komposisi data *training* (pelatihan) dan data *testing* (pengujian) dengan perbandingan data 80:20.

Tabel 4. Hasil Distribusi Decision Tree *One Against All*

Normal	Baik	Kurang Baik
25	3	2

Pie Chart

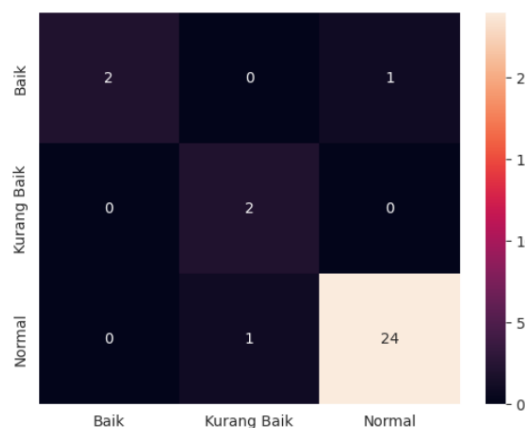


Gambar 5. Distribusi Klasifikasi Decision Tree

Berdasarkan hasil Tabel 4 dan Gambar Grafik 5 diatas digunakan 30 data pelatihan yang dapat disimpulkan bahwa 25 perusahaan termasuk dalam kategori normal atau sebanyak 83%. Dan 3 perusahaan atau 7 persen termasuk kategori baik. Sisanya sebanyak 3 perusahaan atau 10% termasuk dalam kategori kurang baik. Hasil dari algoritma *Decision Tree* mengenai *accuracy*, *F1-score*, *precision*, *recall*, *K-Fold Validation* dan *confusion matrix* berikut ini.

Tabel 5. Hasil Kinerja Decision Tree

Accuracy	Precision	Recall	F1-score	K-Fold Validation
93,33%	87,56%	87,56%	85,33%	90%



Gambar 6. Confusion Matrix Decision Tree

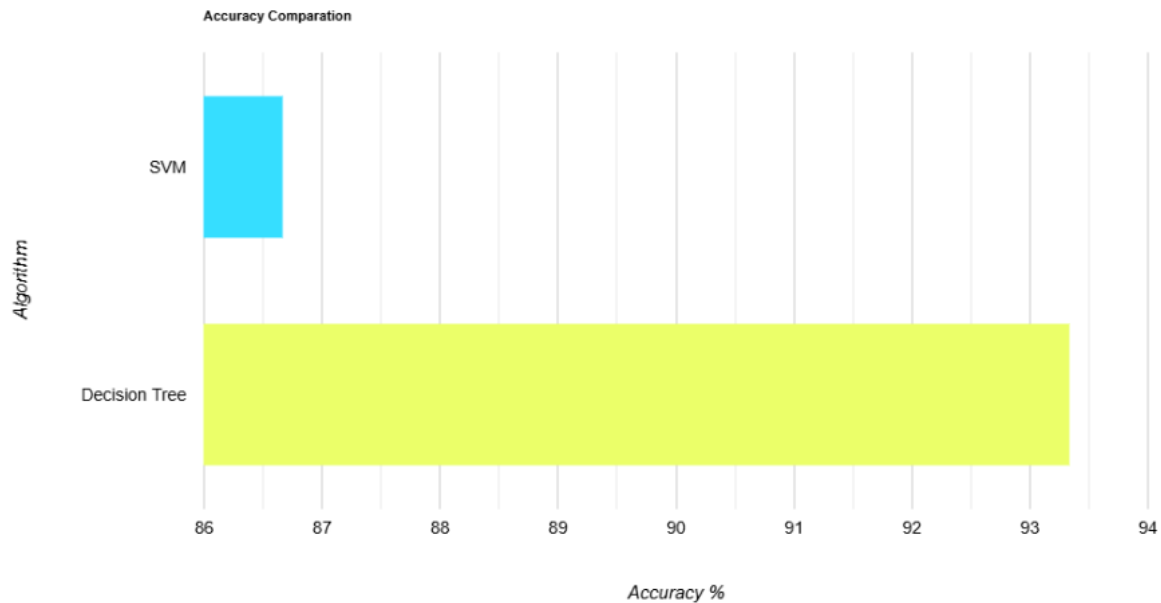
Berdasarkan hasil Tabel 5 dan Gambar 6 setelah dilakukan pengujian menggunakan algoritma *Decision Tree* dan menghasilkan nilai yang meliputi *accuracy*, *precision*, *recall*, *F1-score*, *K-fold validation* dan *confusion matrix* dari *Decision Tree*. Maka hasil dari pengujian tersebut memiliki tingkat nilai *accuracy* sebesar 93,33%, nilai *precision* sebesar 87,56%, nilai *recall* sebesar 87,56%, nilai *F1-score* sebesar 85,33%, dan hasil *accuracy* setelah dilakukannya *cross-validation* dengan k=5 sebesar 90%.

3.6 Comparasion

Penelitian ini menemukan dua tabel perbandingan antara algoritma *Support Vector Machine* dan *Decision Tree*. Hasil dari setiap tabel perbandingan menunjukkan hasil yang berbeda.

Tabel 6. Perbandingan SVM dan Decision Tree

Algoritma	Accuracy	Precision	Recall	F1-score	K-Fold Validation
Support Vector Machine	86,67%	69,56%	75,11%	70,67%	90%
Decision Tree	93,33%	87,56%	87,56%	85,33%	90%



Gambar 7. Perbandingan Akurasi

Berdasarkan hasil perbandingan pada Tabel 6 dan Gambar 7 diatas, menunjukkan bahwa algoritma *Decision Tree* menghasilkan tingkat akurasi yang paling tinggi, sebanyak 93,33%. Namun, tingkat akurasi algoritma *Support Vector Machine* hanya 86,67%. Pada hasil di atas, kita dapat melihat bahwa algoritma *Decision Tree* memiliki nilai *precision* sebesar 87,56%, sementara algoritma *Support Vector Machine* memiliki nilai *precision* sebesar 69,56%. Selain itu, kita dapat melihat bahwa algoritma *Decision Tree* menghasilkan nilai *recall* yang lebih baik, sebesar 87,56%, jika dibandingkan dengan algoritma SVM. Nilai *F1-Score* pada algoritma *Decision Tree* memiliki nilai yang baik, sebesar 85,33% jika dibandingkan dengan algoritma SVM dengan hasil *F1-Score* 70,67%. Pada pengujian validasi menggunakan *cross validation* dengan $k=5$ hasil kinerja dari kedua algoritma tersebut bernilai sama yaitu sebesar 90%. Hal ini mendapatkan nilai estimasi model yang lebih stabil. Maka dari itu peneliti menyimpulkan bahwa untuk melakukan klasifikasi kinerja performa perusahaan, kinerja algoritma *Decision Tree* memiliki nilai akurasi yang lebih tinggi daripada algoritma *Support Vector Machine*. Hal ini ditunjukkan oleh nilai akurasi dan *confusion matrix* yang lebih baik dari algoritma *Support Vector Machine*.

4. KESIMPULAN

Sejauh mana sebuah perusahaan mencapai tujuan bisnisnya dan menghasilkan nilai lebih bagi pemegang saham dan pemangku kepentingan lainnya disebut performa perusahaan. Melakukan klasifikasi adalah salah satu cara untuk mengetahui seberapa baik suatu perusahaan itu menjalankan bisnisnya. Mencari algoritma klasifikasi yang memiliki nilai akurasi tertinggi untuk mengklasifikasikan performa perusahaan adalah tujuan utama dari penelitian ini. Penelitian ini akan mengklasifikasi performa perusahaan berdasarkan rasio keuangan pada laporan keuangan dengan mempertimbangkan perbandingan dua algoritma pembelajaran mesin yaitu *Support Vector Machine* dan *Decision Tree* dengan melakukan pendekatan *One Against All*. *Feature* yang digunakan untuk klasifikasi performa perusahaan ini terdiri dari 3 rasio keuangan yaitu rasio profitabilitas (ROA), likuiditas (CR), dan leverage (DER). Label atau target dalam klasifikasi performa perusahaan ini akan dibagi menjadi 3 kategori yaitu normal, baik, dan kurang baik. Kinerja algoritma *Support Vector Machine* menghasilkan akurasi 86,67%, sedangkan pada algoritma *Decision Tree* menghasilkan akurasi 93,33%. Sehingga algoritma *Decision Tree* menghasilkan hasil yang lebih akurat dibandingkan algoritma *Support Vector Machine* dalam klasifikasi prediksi performa perusahaan berdasarkan rasio keuangan pada laporan keuangan perusahaan. Penelitian ini dilakukan hanya menggunakan beberapa rasio keuangan saja, sehingga untuk penelitian yang akan dikembangkan selanjutnya dapat menggunakan rasio keuangan dan metode evaluasi yang lebih banyak lagi.

REFERENCES

- [1] T. Destiani and R. M. Hendriyani, "Analisis Rasio Keuangan untuk Menilai Kinerja Keuangan Perusahaan," *Al-Kharaj J. Ekon. Keuang. Bisnis Syariah*, vol. 4, no. 1, pp. 33–51, 2021, doi: 10.47467/alkharaj.v4i1.488.
- [2] D. Tambunan, "Investasi Saham di Masa Pandemi COVID-19," *Widya Cipta J. Sekr. dan Manaj.*, vol. 4, no. 2, pp. 117–123, 2020, doi: 10.31294/widyacipta.v4i2.8564.
- [3] Jony, "Sistem Rekomendasi Pemilihan Saham LQ45 Menggunakan Metode Topsis pada Banking," *JISyCS (Journal Inf. Syst. Comput. Sci.*, vol. 1, no. 2, pp. 59–66, 2021, [Online]. Available: <http://www.jurnal.stmikmbcpalembang.ac.id/index.php/JISyCS/article/view/13>.
- [4] E. Giyartiningrum, U. F. Azzahra, H. Prasetyo, and F. Swandari, "Analisis Faktor-Faktor yang Berpengaruh terhadap Return



- Saham Perusahaan Jasa di Bursa Efek Indonesia pada Periode 2018-2021,” *J. Samudra Ekon. dan Bisnis*, vol. 14, no. 2, pp. 331–340, 2023, doi: 10.33059/jseb.v14i2.7028.
- [5] L. R. Harahap, R. Anggraini, E. Ellys, and R. Y. Effendy, “Analisis Rasio Keuangan Terhadap Kinerja Perusahaan Pt Eastparc Hotel, Tbk (Masa Awal Pandemi Covid-19),” *Compet. J. Akunt. dan Keuang.*, vol. 5, no. 1, p. 57, 2021, doi: 10.31000/competitive.v5i1.4050.
- [6] M. Veronica and R. A. Pebriani, “Pengaruh Faktor Fundamental Dan Makro Ekonomi Terhadap Harga Saham Pada Perusahaan Industri Properti Di Bursa Efek Indonesia,” *Islam. Bank. J. Pemikir. dan Pengemb. Perbank. Syariah*, vol. 6, no. 1, pp. 119–138, 2020, doi: 10.36908/isbank.v6i1.155.
- [7] A. L. Ulum, I. Sosial, D. A. N. Humaniora, P. Terhadap, and K. Keuangan, “PENGARUH RASIO LIKUIDITAS, SOLVABILITAS, AKTIVITAS DAN PROFITABILITAS TERHADAP KINERJA KEUANGAN PERUSAHAAN TELEKOMUNIKASI YANG TERDAFTAR DI BURSA EFEK INDONESIA (BEI) PERIODE TAHUN 2016-2020,” *J. Ilmu Sos. dan Hum.*, vol. 8, no. April, pp. 32–44, 2022, doi: <http://dx.doi.org/10.31602/alsh.v8i1.6830>.
- [8] G. D. Desriyunia, D. Wulandhari, D. Puspita, Jasmine, and T. Yulaeli, “Faktor-faktor rasio keuangan meliputi: Rasio likuiditas, rasio solvabilitas, rasio profitabilitas, rasio aktivitas, dan rasio investasi, berpengaruh terhadap kinerja laporan keuangan (literature review manajemen keuangan),” *SAMMAJIVA J. Penelit. Bisnis dan Manaj.*, vol. 1, no. 3, pp. 131–155, 2023, [Online]. Available: <https://doi.org/10.47861/sammajiva.v1i2.356>.
- [9] N. Tambun, M. Mangantar, and V. N. Untu, “Analisis Pengaruh Rasio Likuiditas, Solvabilitas Dan Profitabilitas Terhadap Nilai Perusahaan Pada Perusahaan Telekomunikasi Indonesia Yang Terdaftar Di Bursa Efek Indonesia (Bei) Periode 2017-2021 Analysis of the Effect of Liquidity Ratio, Solvency and P,” *735 J. EMBA*, vol. 10, no. 4, pp. 735–746, 2022.
- [10] R. Tuntun, K. Kusriani, and K. Kusnawi, “Analisis Perbandingan Kinerja Algoritma Klasifikasi dengan Menggunakan Metode K-Fold Cross Validation,” *J. Media Inform. Budidarma*, vol. 6, no. 4, p. 2111, 2022, doi: 10.30865/mib.v6i4.4681.
- [11] Generosa Lukhayu Pritalia, “Analisis Komparatif Algoritme Machine Learning dan Penanganan Imbalanced Data pada Klasifikasi Kualitas Air Layak Minum,” *KONSTELASI Konvergensi Teknol. dan Sist. Inf.*, vol. 2, no. 1, pp. 43–55, 2022, doi: 10.24002/konstelasi.v2i1.5630.
- [12] R. Pakan and E. Mailoa, “Komparasi Kinerja Algoritma Mechine Learning Untuk Mengklasifikasi Penyakit Kanker Payudara,” *J. Ilm. Komput.*, vol. 20, no. 1, pp. 493–503, 2024, doi: 10.35889/progresif.v20i1.1812.
- [13] D. Sепthya et al., “Implementasi Algoritma Decision Tree dan Support Vector Machine untuk Klasifikasi Penyakit Kanker Paru,” *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 3, no. 1, pp. 15–19, 2023, doi: 10.57152/malcom.v3i1.591.
- [14] M. Amirudin and A. D. Wowor, “Analisis Perbandingan Klasifikasi Balita Beresiko Stunting Menggunakan Metode Support Vector Machine dan Decission Tree,” *Conf. Electr. Eng. Informatics, Ind. Technol. Creat. Media*, vol. 3, pp. 1–11, 2023.
- [15] J. Bonardo Junior, R. Rohmat Saedudin, and V. P. Widharta, “Perbandingan Akurasi Algoritma Decision Tree Dan Algoritma Support Vector Machine Pada Penyakit Diabetes,” *e-Proceeding Eng.*, vol. 8, no. 5, pp. 9749–9756, 2021.
- [16] N. G. Ramadhan and A. Khoirunnisa, “Klasifikasi Data Malaria Menggunakan Metode Support Vector Machine,” *J. Media Inform. Budidarma*, vol. 5, no. 4, p. 1580, 2021, doi: 10.30865/mib.v5i4.3347.
- [17] R. S. Tantika, P. Statistika, F. Matematika, and P. Alam, “Klasifikasi pada Data Pasien Penyakit Tiroid dengan Metode SVM,” *Bandung Conf. Ser. Stat.*, vol. 2, no. 2, pp. 159–166, 2022, doi: <https://doi.org/10.29313/bcss.v2i2.3590>.
- [18] D. Ananda and R. R. Suryono, “Analisis Sentimen Publik Terhadap Pengungsi Rohingya di Indonesia dengan Metode Support Vector Machine dan Naïve Bayes,” *J. Media Inform. Budidarma*, vol. 8, no. April, pp. 748–757, 2024, doi: 10.30865/mib.v8i2.7517.
- [19] M. Napiah, R. D. Astuti, and E. K. Pratama, “Komparasi Algoritma Machine Learning untuk Klasifikasi Gejala Coronavirus Disease 19 (Covid-19),” *Comput. Sci.*, vol. 3, no. 2, pp. 78–83, 2023, doi: 10.31294/coscience.v3i2.1984.
- [20] M. H. Wicaksono, M. D. Purbolaksono, and S. Al Faraby, “Perbandingan Algoritma Machine Learning untuk Analisis Sentimen Berbasis Aspek pada Review Female Daily,” *eProceedings Eng.*, vol. 10, no. 3, pp. 3591–3600, 2023.
- [21] F. N. Dhewayani, D. Amelia, D. N. Alifah, B. N. Sari, and M. Jajuli, “Implementasi K-Means Clustering untuk Pengelompokan Daerah Rawan Bencana Kebakaran Menggunakan Model CRISP-DM,” *J. Teknol. dan Inf.*, vol. 12, no. 1, pp. 64–77, 2022, doi: 10.34010/jati.v12i1.6674.
- [22] T. P. Humairah, M. Z. Hakim, and D. S. Abbas, “Pengaruh Profitabilitas, Likuiditas, Dan Leverage Terhadap Nilai Perusahaan,” *Pros. Semin. Nas. Ekon. DAN BISNIS*, vol. 1, no. June, pp. 339–351, 2021, doi: <https://doi.org/10.32528/psneb.v0i0.5186>.