



## Penerapan *Machine Learning* dalam Pengklasifikasian Indeks Saham SRI-Kehati

Muhammad Akbar<sup>1</sup>, Galuh Retno Utami<sup>2</sup>, Bintang Putri Aulia<sup>3</sup>, Robert Kurniawan<sup>4</sup>

<sup>1,3</sup>Program Studi Statistika, Politeknik Statistika STIS, Jakarta, Indonesia

<sup>4</sup>Program Studi Komputasi Statistik, Politeknik Statistika STIS, Jakarta, Indonesia

<sup>1</sup>[212112202@stis.ac.id](mailto:212112202@stis.ac.id)

<sup>3</sup>[212111959@stis.ac.id](mailto:212111959@stis.ac.id)

<sup>4</sup>[robertk@stis.ac.id](mailto:robertk@stis.ac.id)

<sup>2</sup>Program Studi Statistika, Politeknik Statistika STIS, Jakarta, Indonesia

Corresponding author email: <sup>2</sup>[212112068@stis.ac.id](mailto:212112068@stis.ac.id)

**Abstract:** Investors are very concerned about their choices regarding investment in the capital market. One of the considerations for these investment decisions is a new paradigm known as ESG (Environmental, Social, and Governance). In Indonesia, there is only one stock index that applies the principles of Sustainable Responsible Investment (SRI), as well as environmental, social, and governance (ESG) principles, the name is the SRI-Kehati index. The SRI-Kehati Index is a green index published by the KEHATI Foundation in collaboration with the Indonesian Stock Exchange. Therefore, this research aims to identify shares that comply with the SRI-Kehati criteria and provide recommendations for sustainable investment. The data used in this research is secondary data totaling 924 stock companies and comes from the Indonesian Stock Exchange. The research variables used are the SRI-Kehati index as the dependent variable and the type of recording board, stock sector, Syariah stock index, and LQ45 stock index as the independent variables. The method used in this research is Random Over Sampling Examples (ROSE) to overcome data imbalance; random forest, decision tree, naïve Bayes, and support vector machine (SVM) methods for classification. The research results show that the best classification method based on the evaluation measure value is the decision tree method because it has the greatest accuracy, precision, recall, and F1 score.

**Keywords:** SRI-Kehati, Random Forest, Decision Tree, Naïve Bayes, SVM

**Abstrak:** Para investor sangat memperhatikan pilihan mereka tentang investasi di pasar modal. Salah satu pertimbangan keputusan investasi tersebut adalah paradigma baru yang dikenal sebagai ESG (*Environment, Social and Governance*). Satu indeks saham di Indonesia, yaitu indeks SRI-Kehati, menerapkan prinsip *Sustainable Responsible Investment* (SRI), serta prinsip lingkungan, sosial, dan tata Kelola (ESG). Indeks SRI-Kehati adalah indeks hijau yang diterbitkan oleh Yayasan KEHATI yang bekerjasama dengan Bursa Efek Indonesia (BEI). Tujuan dari penelitian ini adalah untuk mengidentifikasi saham yang sesuai dengan kriteria SRI-Kehati serta memberikan rekomendasi untuk investasi berkelanjutan. Adapun data yang dipakai pada penelitian ini adalah data sekunder yang berjumlah 924 perusahaan saham dan berasal dari Bursa Efek Indonesia (BEI). Variabel penelitian yang digunakan adalah indeks SRI-Kehati sebagai variabel dependen dan jenis papan pencatatan, sektor saham, indeks saham syariah serta indeks saham LQ45 sebagai variabel independen. Metode yang digunakan pada penelitian ini adalah *Random Over Sampling Examples* (ROSE) untuk mengatasi ketidakseimbangan data; metode *random forest*, *decision tree*, *naïve bayes* serta *support vector machine* (SVM) untuk klasifikasi. Hasil penelitian menunjukkan bahwa metode klasifikasi terbaik berdasarkan nilai ukuran evaluasi adalah metode *decision tree* karena memiliki akurasi, presisi, *recall* dan *F1-score* terbesar.

**Kata kunci:** SRI-Kehati, *Random Forest*, *Decision Tree*, *Naïve Bayes*, SVM

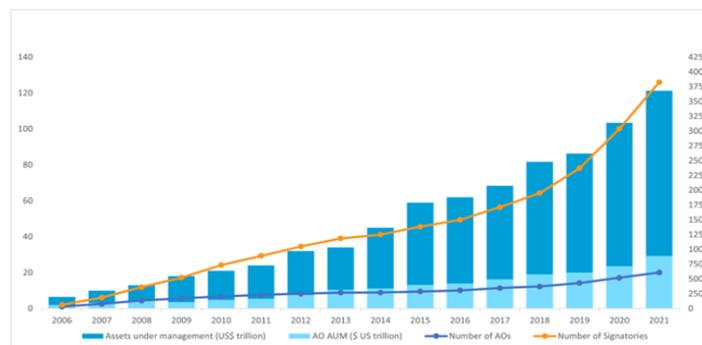
### I. PENDAHULUAN

Para investor, baik individu maupun lembaga keuangan, sangat memperhatikan pilihan mereka tentang investasi di pasar modal. Salah satu pertimbangan dalam proses pengambilan keputusan investasi tersebut telah ditentukan oleh suatu paradigma baru. Paradigma tersebut dikenal dengan istilah ESG yang merupakan singkatan dari *Environment* (Lingkungan), *Social* (Sosial), dan *Governance* (Tata Kelola). Isu lingkungan, sosial, serta kualitas tata kelola perusahaan yang akan diinvestasikan, adalah faktor utama dalam keputusan investasi ESG [1]. Isu lingkungan yang menjadi perhatian saat ini meliputi kelestarian alam yang menjadi sumber penopang kehidupan dalam ketersediaan pangan, udara



yang bersih dan ketersediaan energi. Selain itu, terdapat isu ekonomi sirkular, ketersediaan air bersih di tengah peristiwa perubahan iklim, pemanfaatan metode *fracking* untuk mengoptimalkan pengambilan minyak bumi, dan meminimalisir konsentrasi gas metana yang telah menyumbangkan 25% dalam pemanasan global. Isu sosial yang menjadi perhatian saat ini mengenai dampak dari pandemi *Covid-19*, hak asasi manusia, perbudakan modern dan hak buruh, pekerjaan layak, dan lainnya. Sedangkan topik yang menjadi perhatian dalam isu tata kelola meliputi keadilan pajak, keterlibatan politik yang bertanggung jawab, keamanan siber, perilaku anti korupsi, tujuan perusahaan, nominasi dan pemilihan pemimpin yang dapat mewakili hak para pemegang saham, pelaporan pelanggaran, serta pembayaran eksekutif perusahaan [2].

Dalam mendukung keputusan investor atas kepedulian tanggung jawab sosial dan lingkungan dalam mengambil keputusan investasi ESG, UN (*United Nations*) mengatur dalam enam prinsip investasi bertanggung jawab, umumnya disebut dengan UNPRI (*United Nation Principle Responsibility Investment*). Prinsip-prinsip ini dikembangkan secara independen oleh investor institusional yang bersedia untuk menyusun dan menandatangani prinsip-prinsip tersebut, kemudian ikut menerapkannya. Peningkatan jumlah investor yang menandatangani UNPRI dari tahun ke tahun ditampilkan pada Gambar 1 berikut.



**Gambar 1.** Jumlah Investor Penandatanganan UNPRI  
Sumber : *unpri.org*

Trend peningkatan investor yang ikut menandatangani UNPRI dapat dilihat pada Gambar 1 yang cukup signifikan dari tahun ke tahun.. Hal ini menunjukkan bahwa kepedulian investor terhadap isu ESG dan peluang dalam berkembangnya penerapan UNPRI juga meningkat.

Para investor percaya bahwa masalah lingkungan, sosial, dan tata kelola perusahaan (ESG) dapat memengaruhi keuntungan jangka panjang dan kinerja portofolio investasi. Para investor juga menyadari bahwa penerapan prinsip-prinsip ini dapat lebih menyelaraskan investor dengan tujuan masyarakat yang lebih luas. Dalam prinsip pertama disebutkan bahwa para investor harus memasukkan isu ESG dalam proses pengambilan keputusan investasi. Kedua, investor akan menjadi pemilik aktif dan memasukkan masalah ESG ke dalam kebijakan dan praktik kepemilikan. Ketiga, investor akan mengupayakan pengungkapan yang sesuai tentang masalah ESG oleh entitas tempat berinvestasi. Keempat, investor akan mempromosikan penerimaan dan implementasi prinsip. Kelima, investor akan bekerja sama untuk meningkatkan efektivitas dalam penerapan prinsip. Prinsip terakhir, masing-masing investor akan melaporkan kegiatan dan kemajuan menuju penerapan prinsip-prinsip tersebut. Dengan menerapkan prinsip-prinsip tersebut, UN-PRI mendorong investor untuk menggunakan investasi yang bertanggung jawab untuk meningkatkan return dan mengelola risiko dengan lebih baik, tetapi tidak beroperasi untuk keuntungannya sendiri [3].



Institusi pasar modal yang mengatur negosiasi jual beli saham di Indonesia adalah Bursa Efek Indonesia (BEI). BEI menyediakan publik informasi harga saham yang lengkap melalui media cetak dan elektronik. Indeks saham adalah alat statistik yang digunakan untuk mengetahui pergerakan harga total dari kumpulan saham tertentu. Hingga tahun 2024, 44 indeks saham dicatat di BEI, termasuk indeks SRI-Kehati [4].

Indeks Saham *Sustainable and Responsible Investment* SRI-Kehati merupakan indeks hijau yang diterbitkan oleh Yayasan KEHATI bekerja sama dengan BEI yang diluncurkan pada tanggal 8 Juni 2009, dengan mengacu pada UNPRI. Dalam memilih perusahaan, Indeks Saham SRI-Kehati memiliki standar perusahaan yang menerapkan prinsip *Sustainable Responsible Investment* (SRI), serta prinsip lingkungan, sosial dan tata kelola (ESG). Saat ini, satu-satunya referensi bagi prinsip investasi yang menitikberatkan pada masalah ESG di pasar modal Indonesia adalah Indeks Saham SRI-Kehati. Yayasan KEHATI bertujuan untuk menghubungkan dunia bisnis dan konservasi melalui indeks SRI-Kehati [5].

Dalam melakukan investasi umumnya berlangsung selama bertahun-tahun. Oleh karena itu, dalam menentukan keputusan berinvestasi harus mempertimbangkan perbandingan biaya dan manfaat selama masa pakainya. Manfaat atau keuntungan nilai diharapkan dari investasi harus melebihi nilai sekarang [6]. Manusia adalah individu yang rasional dan menghindari risiko, hal ini juga berlaku bagi investor. Sehingga dalam membuat keputusan berinvestasi, investor akan lebih memilih risiko rendah daripada risiko tinggi serta memiliki tingkat pengembalian (*return*) tertentu [7]. Return indeks di Indonesia mengalami pergerakan yang fluktuatif dari tahun ke tahun.



**Gambar 2.** Return Indeks Saham di Indonesia  
Sumber : TradingView.com

Dalam usaha mendapatkan nilai *return* yang optimal dan memperkirakan risiko dalam membuat keputusan berinvestasi, investor menggunakan strategi dan teori keuangan standar yang berbeda [8]. Berdasarkan Gambar 2 dapat dilihat bahwa sepanjang waktu dari 2010-2024 indeks saham SRI-Kehati dapat mengalahkan *return* indeks paling populer di Indonesia yaitu IDX30, LQ45, dan JII. Ditambah lagi, pada tahun 2022-2024 return indeks saham SRI-Kehati memiliki nilai yang cukup jauh berbeda dengan indeks populer seperti IDX30, LQ45, dan JII. Indeks saham SRI-Kehati menjadi salah satu indeks potensial yang memberikan *return* tinggi.

Berdasarkan latar belakang yang telah dijelaskan, maka tujuan dari penelitian ini adalah untuk mengklasifikasikan saham yang sesuai dengan kriteria SRI-Kehati serta memberikan rekomendasi untuk investasi berkelanjutan. Penelitian sebelumnya mengenai indeks saham SRI-Kehati secara umum menggunakan analisis regresi dan belum banyak ditemukan secara spesifik yang menggunakan penerapan *machine learning* dalam melakukan klasifikasi. Penelitian ini juga melakukan perbandingan



metode *Random Forest*, *Decision Tree*, *Naive Bayes*, dan *Support Vector Machine* (SVM) dalam pengklasifikasiannya.

## II. METODE PENELITIAN

### 2.1 Sumber Data dan Variabel Penelitian

Sumber data pada penelitian ini merupakan data sekunder dari Bursa Efek Indonesia. *Dataset* yang dikumpulkan merupakan perusahaan saham di Indonesia yang terdaftar secara resmi di Bursa Efek Indonesia sebanyak 924 perusahaan. Variabel yang digunakan pada penelitian ini antara lain.

**Tabel 1.** Variabel Penelitian

No	Variabel	Kategori
1.	SRI-Kehati (Y)	1 : Ya 2 : Tidak
2.	Papan Pencatatan	1 : Utama 2 : Pengembangan 3 : Pemantauan Khusus 4 : Ekonomi Baru 5 : Akselerasi
3.	Sektor	1 : Barang Baku 2 : Barang Konsumen Primer 3 : Barang Konsumen Non-Primer 4 : Energi 5 : Keuangan 6 : Kesehatan 7 : Industrial 8 : Infrastruktur 9 : Properti dan Real Estate 10 : Teknologi 11 : Transportasi dan logistik
4.	Syariah	1 : Ya 2 : Tidak
5.	LQ45	1 : Ya 2 : Tidak

### 2.2 Metode Penelitian

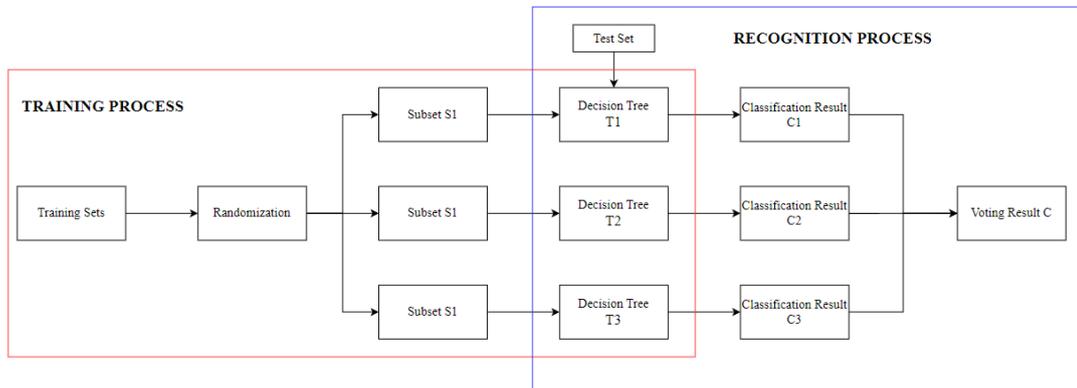
#### 2.2.1 *Random Over Sampling Examples* (ROSE)

ROSE merupakan salah satu teknik untuk mengatasi ketidakseimbangan data akibat terdapat kelas pada dataset yang memiliki jumlah sampel jauh lebih banyak dibandingkan kelas lainnya dengan mengubah sampel menjadi lebih seimbang. Perbedaan ROSE dengan metode over sampling lainnya adalah ROSE menghidupkan kembali efek ketidakseimbangan baik pada fase estimasi model maupun penilaian model dan menyediakan teknik yang lebih halus untuk menemukan kelas minoritas secara efektif dan akurat [9]. Hal tersebut karena pertama, ROSE memilih menggunakan kurva ROC untuk mengukur kinerja klasifikasi dibandingkan akurasi. Selain itu, metode estimasi dilakukan dengan pendekatan *bootstrap* [9], [10]. Dengan kata lain, ROSE menambahkan data pada suatu kelas dengan data bantuan yang semirip mungkin dengan data asli berdasarkan pendekatan *smoothed bootstrap* [11].

#### 2.2.2 *Random Forest*

*Random forest* merupakan salah satu metode klasifikasi populer yang terdiri dari banyak pohon keputusan (*decision tree*) yang independen satu sama lain. *Random forest* dapat mengklasifikasikan data tidak seimbang dalam jumlah besar dengan hasil yang baik serta waktu pemrosesan yang relatif

cepat [12]. Selain itu, *random forest* juga memiliki kelebihan lain seperti besar *error* relatif rendah, hasil klasifikasi yang baik, mampu mengatasi data training yang berjumlah besar secara efisien dan efektif dalam mengatasi *missing* data [13].



**Gambar 3.** Proses Klasifikasi *Random Forest*  
Sumber : Parmar, 2019

Gambar 3 merupakan proses klasifikasi *random forest* [12]. Cara pengklasifikasian pada *random forest* dilakukan dengan cara mengadopsi pendekatan ansambel dari bermacam-macam pohon keputusan melalui kemunculan paling banyak untuk mencapai keputusan akhir [14].

### 2.2.3 Decision Tree

*Decision tree* (pohon keputusan) merupakan metode data mining yang paling umum digunakan. Metode ini menggunakan *flowchart* berupa pohon keputusan yang mengkategorikan sekumpulan data ke dalam kelas atau kategori tertentu [15]. *Decision tree* merupakan metode pengklasifikasian dengan algoritma non-parametrik yang mampu menangani data berukuran besar dan rumit secara efisien tanpa menerapkan struktur parametrik yang rumit [16], [17]. Kelebihan lain dari *decision tree* adalah menghindari timbulnya permasalahan dan tetap menjaga kualitas keputusan yang dihasilkan [17].

Data dalam *decision tree* berupa tabel dengan atribut dan *record*. Atribut terbagi menjadi dua yaitu sebagai variabel target dan variabel prediktor. *Decision tree* akan memprediksi variabel target dengan menggunakan nilai dari sekumpulan variabel prediktor [15], [17]. *Decision tree* terdiri dari tiga node yaitu node akar, node internal dan node daun. Pada node akar dan node internal dilakukan pengklasifikasian atribut dan hasil keputusan tertulis pada node daun [15].

Penelitian ini menggunakan menggunakan algoritma C4.5, yang merupakan pengembangan dari algoritma ID3 dengan peningkatan berupa penanganan atribut-atribut numerik, *missing value* dan *noise* pada dataset, dan aturan-aturan yang dihasilkan dari model pohon yang terbentuk. Algoritma C4.5 membentuk struktur pohon keputusan dengan simpul yang menggambarkan atribut dan kelas, dan setiap cabang menggambarkan hasil dari atribut yang diuji [18]. Langkah kerja algoritma C4.5 yaitu menghitung nilai *entropy*, nilai *information gain*, nilai *split info*, dan nilai *gain ratio* [19].

### 2.2.4 Naive Bayes

*Naive Bayes* merupakan metode klasifikasi yang sederhana dengan asumsi seluruh atribut bersifat independen [20]. Metode ini mengklasifikasikan dataset dengan menghitung peluang yang dilakukan dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset [21]. Selain itu, *naive bayes* juga memiliki keunggulan yaitu hanya membutuhkan data *training* berukuran kecil dalam

proses estimasi parameter pada saat pengklasifikasian [22]. Secara umum, bentuk persamaan *naive bayes* sebagai berikut [20].

$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(X)} \quad (1)$$

dimana,

$P(H|X)$  = peluang H jika X telah terjadi

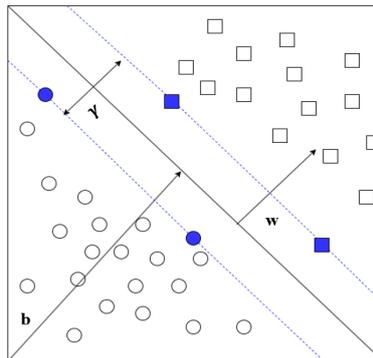
$P(X|H)$  = peluang X jika H telah terjadi

$P(X)$  = peluang kejadian X

$P(H)$  = peluang kejadian H

### 2.2.5 Support Vector Machine (SVM)

*Support Vector Machine* (SVM) merupakan metode yang digunakan untuk klasifikasi ataupun prediksi. SVM bekerja dengan mengidentifikasi *hyperplane* atau jarak pemisah paling optimal (margin maksimum) antar kelas atau kategori yang berbeda. SVM dapat berbentuk linier dan non-linier. Pada SVM linier digambarkan dengan jarak pemisah berupa garis lurus yang memisahkan dua kelas atau kategori (Gambar 4) [23], [24]. *Support vector* adalah data yang paling dekat dengan garis pembatas antar dua kelas atau kategori [23].



**Gambar 4.** *Hyperplane* yang memisahkan dua kelas; lingkaran dan kotak masing-masing mewakili titik data terdekat (*support vector*) -1 dan +1  
Sumber : Mammone, 200

Pada Gambar 4, terdapat garis pembatas antar kelas yang dirumuskan sebagai berikut:

$$w \cdot x_i + b \leq -1 \quad (2)$$

Data  $x_i$  memenuhi pertidaksamaan diatas maka termasuk dalam kelas negatif.

$$w \cdot x_i + b \geq 1 \quad (3)$$

Data  $x_i$  memenuhi pertidaksamaan diatas maka termasuk dalam kelas positif.

dimana,

$$w = \text{vektor pembobot berukuran dihitung dengan } w = \sum_{i=1}^m \alpha_i y_i x_i \quad (4)$$

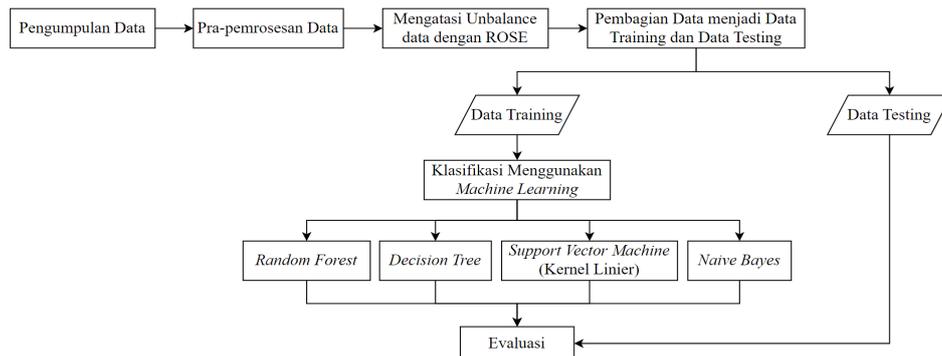
$x_i$  = variabel prediktor

$$b = \text{nilai bias (menentukan posisi } hyperplane) \text{ dihitung dengan } b = y_i - \sum_{j=1}^m \alpha_j y_j (x_j \cdot x_i) \quad (5)$$

Metode ini mengasumsikan bahwa distribusi antar kelas atau kategori seimbang. Hal tersebut dikarenakan apabila terdapat kelas atau kategori yang berisi lebih banyak dapat berdampak negatif pada kinerja pengklasifikasian [23].

### 2.3 Tahapan Penelitian

Penelitian ini melakukan pengklasifikasian dengan beberapa metode klasifikasi yaitu *random forest*, *decision tree*, *naive bayes* dan *support vector machine (SVM)* dengan bantuan *software R*. Dataset diklasifikasikan menjadi dua kategori yaitu termasuk saham SRI-Kehati dan bukan saham SRI-Kehati.

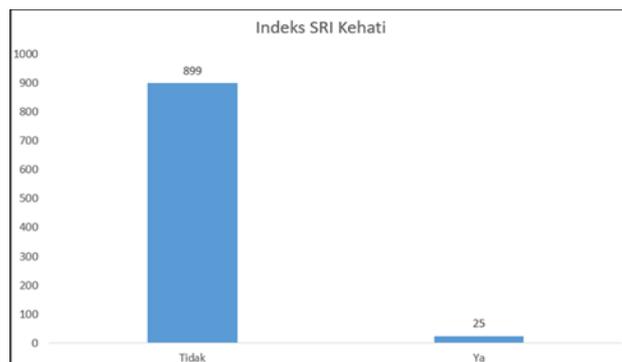


Gambar 5. Tahapan analisis

## III. HASIL DAN PEMBAHASAN

### 3.1 Gambaran Umum Seluruh Saham yang Ada di Bursa Efek Indonesia

Berawal dari dibangunnya Bursa Efek Jakarta (BEJ) pada tahun 1977 yang sekarang sudah berganti menjadi Bursa Efek Indonesia. Sejak awal berdiri dengan banyaknya perusahaan yang melakukan *Initial Public Offering (IPO)* serta beberapa saham dilakukan *delisting*, saat ini jumlah perusahaan terbuka yang tercatat ada di Bursa Efek Indonesia terdapat 924 saham. BEI menerbitkan beberapa indeks saham yang berisi sekumpulan saham berdasarkan kriteria dan evaluasi tertentu, salah satunya indeks SRI-KEHATI yang berisi kumpulan perusahaan dengan model bisnis berkelanjutan, serta memiliki kesadaran terhadap lingkungan hidup, sosial, dan tata kelola perusahaan yang baik.

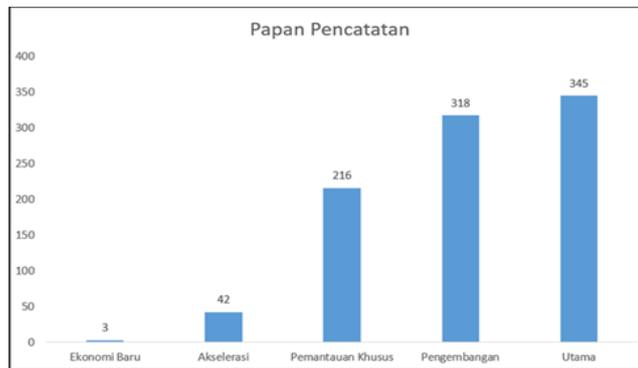


Gambar 6. Frekuensi Saham Indeks SRI Kehati

Indeks SRI-Kehati mempunyai peran yang strategis dalam mewujudkan komitmen yang kuat pada prinsip-prinsip ESG yang memperhatikan aspek berkelanjutan yaitu sosial lingkungan dan juga tata kelola perusahaan yang baik. Indeks ini memperhatikan tren global yang bergerak menuju ekonomi hijau dan keberlanjutan, sehingga SRI-Kehati berpotensi menjadi semakin relevan di masa depan. Perusahaan-perusahaan yang tidak termasuk dalam indeks ini mungkin perlu menilai kembali kebijakan dan praktik bisnis mereka untuk memenuhi kriteria ESG, sehingga dapat menarik lebih banyak investor yang sadar lingkungan dan sosial. Indeks SRI-Kehati tidak hanya berfungsi sebagai indikator

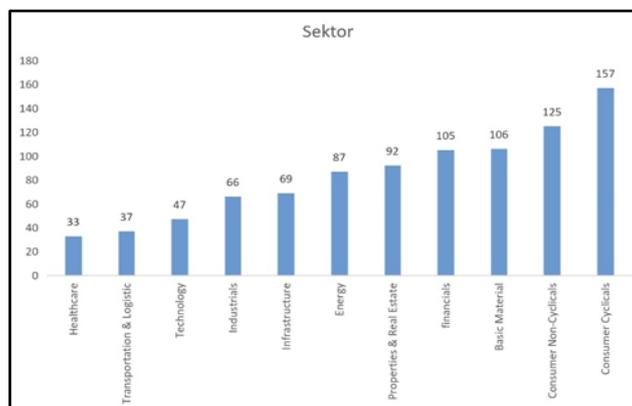


keberlanjutan saat ini tetapi juga sebagai pendorong perubahan positif dalam praktik bisnis di Indonesia. Praktik bisnis perusahaan juga dapat dinilai oleh BEI berdasarkan papan pencatatan yang dibagi menjadi Utama, Akselerasi, Ekonomi Baru, Pengembangan, dan Pemantauan Khusus.



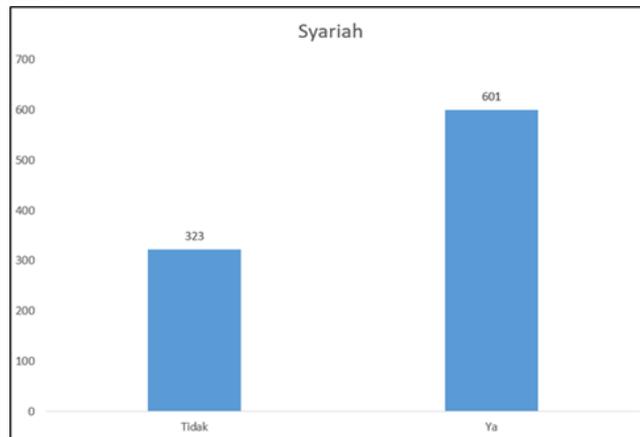
**Gambar 7.** Frekuensi Papan Pencatatan Saham

Papan pencatatan berkaitan dengan hal-hal fundamental dari masing-masing perusahaan. Pemberian papan pencatatan oleh Bursa Efek Indonesia didasarkan pada masa operasional, laba usaha, laporan keuangan yang teraudit, ukuran keuangan perusahaan. Saat ini papan pencatatan terbanyak adalah utama dengan jumlah 345 saham, sedangkan kategori paling sedikit adalah papan pencatatan ekonomi baru. Selain papan pencatatan, hal penting untuk menilai perusahaan yang memiliki prinsip ESG dapat dilihat berdasarkan sektor.



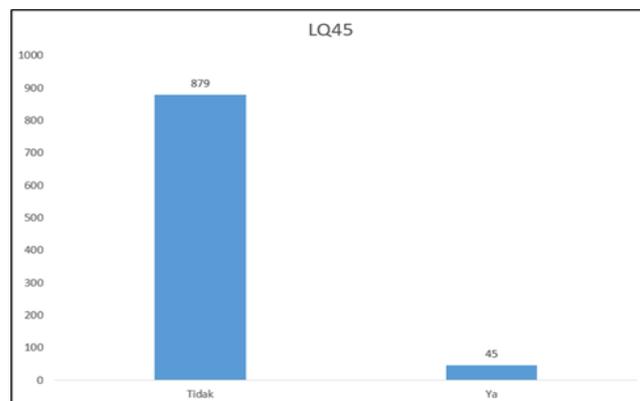
**Gambar 8.** Frekuensi Sektor Saham

Sektor merupakan hal paling esensial dari sebuah investasi hijau karena ini merupakan kegiatan utama yang dilakukan perusahaan apakah itu ramah lingkungan atau tidak. Sektor konsumen siklikal, energi, industrial umumnya tidak memenuhi kriteria *green investment* [25]. Sektor perusahaan terbuka di Indonesia yang paling banyak adalah *consumer cyclicals* seperti *Ace Hardware Indonesia (ACES)*. Sedangkan sektor yang paling sedikit adalah *Healthcare* seperti *Kalbe Farma (KLBF)*.



**Gambar 9.** Frekuensi Saham Syariah

Saham syariah di Indonesia berjumlah 601 perusahaan dengan kriteria-kriteria yang memenuhi syariat Islam. Saham-saham ini tercantum dalam *Indonesia Sharia Stock Index (ISSI)* yang masuk dalam daftar efek syariah yang diterbitkan oleh OJK dan tercatat di papan utama dan papan pengembangan. Indeks ini akan diseleksi ulang sebanyak dua kali setahun di bulan Mei dan November. Selain kategori syariah, penting juga untuk menilai likuiditas volume perdagangan saham. Terdapat indeks yang berisi saham-saham dengan likuiditas besar yaitu LQ45.

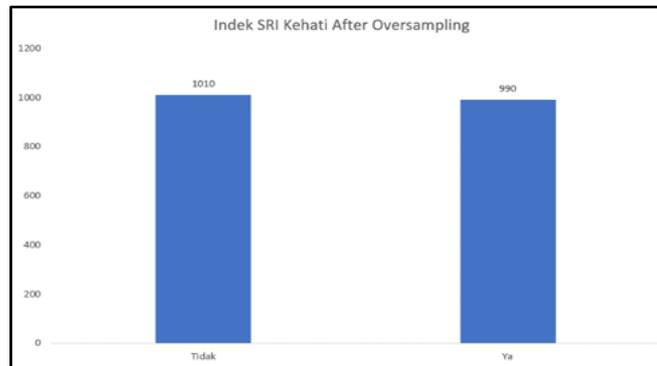


**Gambar 10.** Frekuensi Saham LQ45

Sesuai dengan namanya, Indeks LQ45 berisi 45 saham yang dipilih berdasarkan likuiditas tertinggi dan kapitalisasi pasar terbesar serta kriteria lainnya. Hal ini berarti 45 saham di dalamnya mempunyai volume perdagangan yang tinggi jika dibandingkan dengan 879 saham lainnya. Likuiditas pasar yang tinggi menandakan minat para investor yang tinggi dan membuat para pemegang saham merasa aman karena sebaran penawaran dan permintaan yang seimbang [26].

### 3.2 Oversampling data

Berdasarkan gambar 6, terjadi *imbalance* pada data kelas SRI-Kehati. Efek dari *imbalance* data dalam klasifikasi adalah ketidakcukupan kelas minoritas pada data training serta akurasi yang menyestakan [27]. Salah satu metode yang umum digunakan untuk *imbalance* data adalah *Random Over Sampling Examples (ROSE)*. ROSE dilakukan dengan memilih sampel kelas minoritas secara acak lalu dilakukan duplikasi [28]. Oversampling menyebabkan peningkatan signifikan pada kelas minoritas sehingga data menjadi seimbang. Maka dari itu model klasifikasi tidak akan terjadi masalah *imbalance* data.



Gambar 11. Frekuensi Saham SRI Kehati Setelah Oversampling

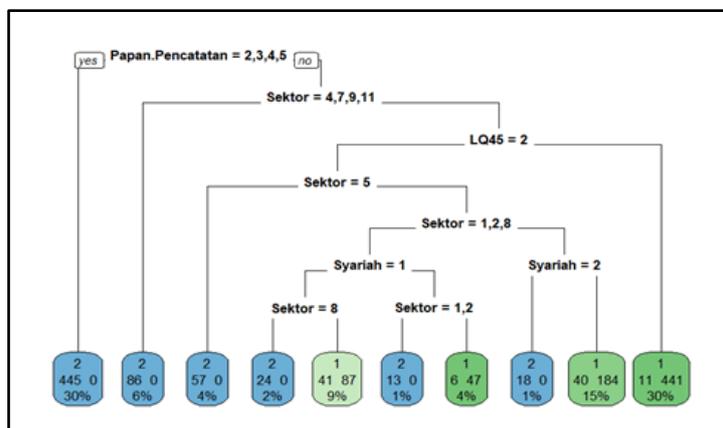
Sebelum dilakukan klasifikasi data yang telah dilakukan *oversampling* akan dibagi menjadi data *training* sebesar 75% dan data *testing* sebesar 25%. Model klasifikasi yang akan digunakan adalah *Random Forest*, *Decision Tree*, *Naive Bayes*, dan *SVM*.

### 3.3 Model Klasifikasi Terbaik

Tabel 2. Perbandingan model klasifikasi

Metode	Akurasi	Presisi	Recall	F1
<i>Random Forest</i>	91,2	0,852	1	0,920
<b><i>Decision Tree</i></b>	<b>92,2</b>	<b>0,855</b>	<b>1</b>	<b>0,922</b>
<i>Naive Bayes</i>	85	0,802	0,896	0,847
<i>SVM</i>	85	0,802	0,896	0,847

Berdasarkan perbandingan model dengan metode *random forest*, *decision tree*, *naive bayes*, dan *SVM* yang dilihat dari ukuran evaluasi model menunjukkan model *decision tree* merupakan model klasifikasi terbaik. Hal ini dapat dilihat dari nilai akurasi, presisi, *recall*, dan F1 Score yang paling besar diantara model lainnya. *Decision tree* akan menghasilkan sebuah pohon yang berisi rule model. Hasil dari pohon ini menunjukkan papan pencatatan sebagai akar dari pohonnya. Apabila papan pencatatan adalah kategori selain 1 maka saham tersebut memiliki label 2 atau bukan termasuk SRI-Kehati. Jika papan pencatatan adalah 1 dan sektor adalah 4,7,9,11 maka saham tersebut juga memiliki label 2 dan seterusnya. Alur pembentukan model *decision tree* dapat dilihat pada gambar berikut.



Gambar 12. Decision Tree Rules



**Tabel 3.** *Confusion Matrix*

Prediction	References	
	Ya	Tidak
Ya	231	39
Tidak	0	230

Berdasarkan *confusion matrix* diatas menunjukkan bahwa sebanyak 231 label “Ya” pada data training dapat diprediksi dengan benar oleh model dan juga label “Tidak” sebanyak 230 dapat diprediksi dengan benar. Sedangkan terdapat 39 data dengan label “Tidak” yang di prediksi salah oleh model. Prediksi yang dilakukan model *decision tree* pada data *testing* menghasilkan akurasi 92,2% yang berarti 92,2% data diprediksi benar oleh model. Lalu nilai *recall*-nya menunjukkan nilai 1 yang berarti tidak menunjukkan adanya *false positive*. Hal ini juga sejalan dengan penelitian Misra et al (2018) dimana penggunaan machine learning lebih baik untuk memprediksi pasar saham dibanding metode lainnya dimana metode SVM mempunyai hasil akurasi yang sangat baik [29]. Begitupun prediksi harga saham yang dilakukan oleh Ravikumar dan Saraf (2020) dimana metode machine learning lebih baik dibanding dengan regresi maupun analisis runtun waktu [30].

#### IV. KESIMPULAN

Berdasarkan hasil penelitian yang telah dijelaskan, maka dapat diambil kesimpulan sebagai berikut:

1. Investasi hijau menjadi paradigma baru dalam menentukan seorang investor untuk memilih jenis aset yang akan mereka investasikan.
2. Klasifikasi saham ke dalam indeks SRI-kehati dengan berbagai metode menghasilkan akurasi 91,2% dengan metode *random forest*, akurasi 92,2% dengan *decision tree*, akurasi 85% dengan metode *naive bayes* dan SVM.
3. Metode klasifikasi terbaik berdasarkan nilai ukuran evaluasi adalah *decision tree* karena memiliki akurasi, presisi, *recall*, dan *F1-score* terbesar.

Saran kepada penelitian selanjutnya adalah melakukan eksplorasi mendalam untuk variabel-variabel yang belum tercakup ke dalam model. Selain itu juga dapat menggunakan metode klasifikasi yang lebih *advance* seperti *deep learning*.

#### UCAPAN TERIMA KASIH

Terima Kasih kami ucapkan kepada Bapak Robert Kurniawan, M.Si., atas bantuan beliau dalam pembimbingan dan penyusunan naskah jurnal ini dan Bursa Efek Indonesia atas penyedia data saham.

#### REFERENSI

1. G. Clark, A. Feiner, and M. Viehs, “From the Stockholder to the Stakeholder,” *Univerisy of Oxford*, no. March, p. 58, 2015, [Online]. Available: [https://arabesque.com/research/From\\_the\\_stockholder\\_to\\_the\\_stakeholder\\_web.pdf](https://arabesque.com/research/From_the_stockholder_to_the_stakeholder_web.pdf).
2. UN PRI, “Environmental , social and governance issues The PRI works with its signatories to identify key environmental , social and governance ( ESG ) issues in the market . It produces guidance materials , organises events and facilitates initiatives to address t,” 2024. <https://www.unpri.org/sustainability-issues/environmental-social-and-governance-issues> (accessed May 30, 2024).
3. UN PRI, “What is the PRI? About the PRI,,” 2022. <https://www.unpri.org/about-us/about-the-pri> (accessed May 30, 2024).
4. IDX, “Indeks,” 2024. <https://idx.co.id/id/produk/indeks/> (accessed Jun. 03, 2024).
5. Yayasan Kehati, “KEHATI.sri-kehati.pdf,” 2024. <https://kehati.or.id/indeks-sri-kehati/> (accessed Jun. 03, 2024).
6. A. C. Harberger, G. P. Jenkins, and C.-Y. Kuo, “Cost-Benefit Analysis for Investment Decisions: Chapter



- 8,” *Queen’s Univ.*, no. October, p. 39, 2007.
7. H. Markowitz, “Portfolio Selecion,” *J. Finance*, vol. 7, no. 1, pp. 77–91, 1952, [Online]. Available: <http://www.jstor.org/stable/2975974>.
8. M. Arora and S. Kumari, “Risk Taking in Financial Decisions as a Function of Age, Gender: Mediating Role of Loss Aversion and Regret,” *Int. J. Appl. Psychol.*, vol. 5, no. 4, pp. 83–89, 2015, doi: 10.5923/j.ijap.20150504.01.
9. J. Zhang and L. Chen, “Clustering-based Undersampling with Random Over Sampling Examples and Support Vector Machine for Imbalanced Classification of Breast Cancer Diagnosis,” *Comput. Assist. Surg.*, vol. 24, no. S2, pp. 62–72, 2019, doi: <https://doi.org/10.1080/24699322.2019.1649074>.
10. N. A. Selamat, A. Abdullah, and N. Mat Diah, “Association features of smote and rose for drug addiction relapse risk,” *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 34, no. 9, pp. 7710–7719, 2022, doi: 10.1016/j.jksuci.2022.06.012.
11. S. DEMİR and E. K. ŞAHİN, “Evaluation of Oversampling Methods (OVER, SMOTE, and ROSE) in Classifying Soil Liquefaction Dataset based on SVM, RF, and Naïve Bayes,” *Eur. J. Sci. Technol.*, no. 34, pp. 142–147, 2022, doi: 10.31590/ejosat.1077867.
12. A. Parmar, R. Katariya, and V. Patel, “A Review on Random Forest: An Ensemble Classifier,” *Lect. Notes Data Eng. Commun. Technol.*, vol. 26, pp. 758–763, 2019, doi: 10.1007/978-3-030-03146-6\_86.
13. L. Breiman, “RFRSF: Employee Turnover Prediction Based on Random Forests and Survival Analysis,” *Kluwer Acad. Publ.*, vol. 45, pp. 5–32, 2001, doi: 10.1007/978-3-030-62008-0\_35.
14. C. Yoo, D. Han, J. Im, and B. Bechtel, “Comparison between convolutional neural networks and random forest for local climate zone classification in mega urban areas using Landsat images,” *ISPRS J. Photogramm. Remote Sens.*, vol. 157, no. February, pp. 155–170, 2019, doi: 10.1016/j.isprsjprs.2019.09.009.
15. Z. Yu, F. Haghghat, B. C. M. Fung, and H. Yoshino, “A decision tree method for building energy demand modeling,” *Energy Build.*, vol. 42, no. 10, pp. 1637–1646, 2010, doi: 10.1016/j.enbuild.2010.04.006.
16. Y. Y. Song and Y. LU, “Decision tree methods: applications for classification and prediction,” *Shanghai Arch. Psychiatry*, vol. 27, no. 2, pp. 130–135, 2015, doi: 10.11919/j.issn.1002-0829.215044.
17. Haudi, *Teknik Pengambilan Keputusan*, vol. 3, no. 2. Solok: Insan Cendekia Mandiri, 2015.
18. I. Sutoyo, “Implementasi Algoritma Decision Tree Untuk Klasifikasi Produk Laris,” *J. Ilm. Ilmu Komput.*, vol. 7, no. 2, pp. 45–51, 2021, doi: 10.35329/jiik.v7i2.203.
19. D. A. H. D. Larasati and T. Sutrisno, “Tourism Site Recommendation in Jakarta Using Decision Tree Method Based on Web Review,” *SSRN Electron. J.*, 2018, doi: 10.2139/ssrn.3268964.
20. T. R. Patil and S. S. Sherekar, “Performance Analysis of ANN and Naive Bayes Classification Algorithm for Data Classification,” *J. Comput. Sci. Appl*, vol. 6, no. 2, p. 256, 2013, doi: 10.18201/ijjsae.2019252786.
21. I. Rish, “An ab initio study of RbO, CsO and FrO (X2Σ<sup>+</sup>; A2Π) and their cations (X3Σ<sup>-</sup>; A3Π),” *IJCAI*, vol. 3, no. 22, pp. 41–46, 2001, doi: 10.1039/b104835j.
22. G. Parthiban, A. Rajesh, and S. K. Srivatsa, “Diagnosis of Heart Disease for Diabetic Patients using Naive Bayes Method,” *Int. J. Comput. Appl.*, vol. 24, no. 3, pp. 7–11, 2011, doi: 10.5120/2933-3887.
23. D. A. Pisner and D. M. Schnyer, *Support vector machine*. Elsevier Inc., 2019.
24. A. Mammone, M. Turchi, and N. Cristianini, “Support Vector Machines ( SVM ) Support Vector Machines ( SVM ),” *John Wiley Sons, Inc*, vol. 1, pp. 283–289, 2009, doi: 10.1002/wics.049.
25. F. A. Bergederick, Laura; Maheshwari, Aditi; Ugaz, “A Bottom-up Approach to Track Existing Flows,” *International Finance Corporation (IFC)*, Washington, D.C, p. 47, 2016.
26. A. Alaoui Mdaghri, A. Raghibi, C. N. Thanh, and L. Oubdi, “Stock market liquidity, the great lockdown and the COVID-19 global pandemic nexus in MENA countries,” *Rev. Behav. Financ.*, vol. 13, no. 1, pp. 51–68, 2021, doi: 10.1108/RBF-06-2020-0132.
27. A. Somasundaram and U. S. Reddy, “Data Imbalance: Effects and Solutions for Classification of Large and Highly Imbalanced Data,” *Proc. 1st Int. Conf. Res. Eng. Comput. Technol. (ICRECT 2016)*, no. November, pp. 28–34, 2016.
28. A. Moreo, A. Esuli, and F. Sebastiani, “Distributional random oversampling for imbalanced text classification,” *SIGIR 2016 - Proc. 39th Int. ACM SIGIR Conf. Res. Dev. Inf. Retr.*, pp. 805–808, 2016, doi: 10.1145/2911451.2914722.
29. M. Misra, A. P. Yadav, and H. Kaur, “Stock Market Prediction using Machine Learning Algorithms: A Classification Study,” *2018 Int. Conf. Recent Innov. Electr. Electron. Commun. Eng. ICRIEEECE 2018*, pp. 2475–2478, 2018, doi: 10.1109/ICRIEECE44171.2018.9009178.



Seminar Nasional Sains Data 2024 (SENADA 2024)  
UPN “Veteran” Jawa Timur

E-ISSN 2808-5841

P-ISSN 2808-7283

30. S. Ravikumar and P. Saraf, “Prediction of stock prices using machine learning (regression, classification) Algorithms,” *2020 Int. Conf. Emerg. Technol. INCET 2020*, pp. 5–7, 2020, doi: 10.1109/INCET49848.2020.9154061.