TANAH GOYANG
JURNAL GEOSAINS

Vol. 2 No. 1, Mei 2024
e-ISSN: 3026-3263

KLASIFIKASI BATUAN BEKU BERDASARKAN DATA GEOKIMIA MENGGUNAKAN ALGORITMA *RANDOM FOREST CLASSIFIER*

CLASSIFICATION OF IGNEOUS ROCKS BASED ON GEOCHEMISTRY DATA USING THE RANDOM FOREST

Muh. Riswan Anas Sukri[1*], Abdi Pangestu[2]

[1,2] Universitas Pattimura; Jalan. Ir. M. Putuhena, Kampus Poka Ambon; (0911)3684030

muh.riswan.as@gmail.com*

Kata Kunci:
Algoritma
Geokimia
Machine Learning
Random Forest

Abstrak.

Identifikasi dan klasifikasi batuan berdasarkan karakteristik visual batuan adalah proses yang subjektif dan menggunakan metode yang sama dapat menghasilkan hasil yang berbeda. Pengembangan *machine learning* telah membuka cara baru untuk mengklasifikasikan batuan. Penelitian ini akan dilakukan untuk mengklasifikasikan batuan beku berdasarkan data geokimia menggunakan algoritma *Random Forest Classifier*. *Random forest* adalah algoritma *machine learning* yang menggunakan kombinasi pohon keputusan untuk membuat prediksi yang akurat guna menentukan cara yang lebih tepat dalam memproses data. Model menunjukkan performa yang cukup baik dengan akurasi 89.4% pada data *test*. Pada data *test*, nilai *precision* berkisar antara 0.75 hingga 1.00, *recall* antara 0.80 hingga 1.00, dan *f1-score* antara 0.78 hingga 0.98. Variabel paling penting dalam model klasifikasi batuan ini adalah SIO2_WT%, dengan penurunan skor rata-rata terbesar sekitar 0.30, diikuti oleh MNO_WT% dan FEOT_WT%. Variabel lain memiliki penurunan skor rata-rata yang lebih kecil, menunjukkan kontribusi yang lebih rendah.

Keywords:
Algorithm
Geochemistry
Machine Learning
Random Forest

Abstract.

Identification and classification of rocks based on their visual characteristics is a subjective process, and using the same method can yield different results. The development of machine learning has opened new ways to classify rocks. This research aims to classify igneous rocks based on geochemical data using the Random Forest Classifier algorithm. Random forest is a machine learning algorithm that uses a combination of decision trees to make accurate predictions, providing a more precise way to process data. The model shows good performance with an accuracy of 89.4% on the test data. On the test data, the precision values range from 0.75 to 1.00, recall from 0.80 to 1.00, and the F1-score from 0.78 to 0.98. The most important variable in this rock classification model is SIO2_WT%, with the largest average score decrease of about 0.30, followed by MNO_WT% and FEOT_WT%. Other variables have smaller average score decreases, indicating lower contributions.

1. PENDAHULUAN

Batuan beku terdiri dari senyawa-senyawa kimia yang membentuk mineral, dan mineral-mineral ini kemudian menyusun batuan tersebut. Salah satu cara untuk mengelompokkan batuan beku adalah dengan menggunakan senyawa oksidanya, seperti SiO_2, TiO_2, Al_2O_3, Fe_2O_3, FeO, MnO, MgO, CaO, Na_2O, K_2O, dan P_2O_5 (Kantu, 2022). Persentase senyawa-senyawa ini dapat mencerminkan jenis batuan, lingkungan pembentukannya, dan karakteristik lainnya. Sifat dan jenis batuan beku juga dapat ditentukan berdasarkan kandungan SiO_2 (Williams *et al.*, 1982).

Para ahli geologi umumnya mengelompokkan jenis-jenis batuan berdasarkan ciri-ciri visual seperti mineralogi, warna, dan tekstur. Pada batuan yang tampak serupa, mereka memanfaatkan data lain, seperti data geokimia (Houshmand *et al.*, 2022). Karakteristik visual merupakan indikator dari atribut kimia dan mineralogi batuan, serta proses pembentukannya. Karakteristik-karakteristik ini sangat penting untuk prediksi jenis batuan yang akurat (Ran *et al.*, 2019). Namun, menurut Li dan Li (2013), identifikasi dan klasifikasi litologi berdasarkan karakteristik visual batuan adalah proses yang subjektif dan menggunakan metode yang sama dapat menghasilkan hasil yang berbeda.

Pengembangan *machine learning* telah membuka cara baru untuk mengklasifikasikan batuan. *Machine learning* mampu menganalisis data dalam jumlah besar untuk secara otomatis mengidentifikasi karakteristik batuan. Hal ini berbeda dengan penilaian manusia dan observasi kualitatif, yang dapat dipengaruhi oleh bias subjektif dan menghasilkan hasil yang tidak konsisten. Selain memberikan pendekatan yang konsisten dan dapat diskalakan, *machine learning* dapat meningkatkan keakuratan klasifikasi (Niu *et al.*, 2024).

Klasifikasi digunakan untuk menemukan model fungsi dan mendeskripsikan data ke kelas-kelas berdasarkan data di masa lampau. Data yang telah dikumpulkan akan dipelajari dan dianalisis hubungannya sesuai dengan label atau target yang telah ditentukan (Mangalathu dan Burton, 2019). *Random forest* adalah algoritma *machine learning* yang menggunakan kombinasi pohon keputusan untuk membuat prediksi yang akurat guna menentukan cara yang lebih tepat dalam memproses data. Kelebihan *Random Forest* adalah kemampuannya untuk menangani kumpulan data yang besar dengan banyak fitur beragam, sehingga mampu mengolah data dengan efektif serta mengatasi masalah *overfitting* yang sering terjadi pada pohon keputusan tunggal. Selain itu, *Random Forest* juga mampu menjaga stabilitas kinerja yang tinggi (Siregar *et al.*, 2023).

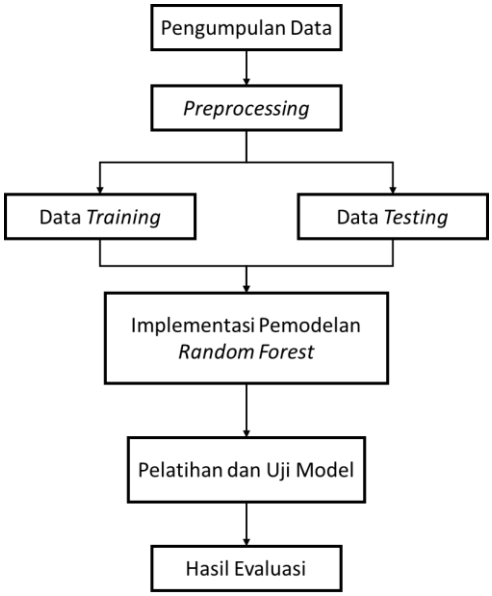
Beberapa penelitian terkait implementasi algoritma klasifikasi yang dilakukan peneliti terdahulu. Klasifikasi dilakukan menggunakan *machine learning supervised* untuk menentukan akurasi terbaik dengan lima jenis metode, antara lain *K-Nearest Neighbors* (K-NN), *Support Vector Machine* (SVM), *Decision Tree* (DT), *Random Forest* (RF), dan *Multi-layer Perceptron* (MLP). Pane dan Sihombing (2020), menemukan bahwa metode SVM memberikan akurasi terbaik pada data SWIR, sedangkan metode MLP unggul pada data TIR untuk klasifikasi mineral batuan. Selain itu, model terbaik diimplementasikan untuk mengidentifikasi empat jenis batuan plutonik dari yang paling gelap hingga yang paling terang: gabro, diorit, granodiorit, dan granit. Menurut Alférez *et al.* (2021), model *K-Nearest Neighbors* memberikan hasil terbaik dengan akurasi, presisi, *recall*, dan *F-score* mencapai 93%. Penelitian lainnya oleh Bamford *et al.* (2021), menunjukkan bahwa algoritma *machine learning* semakin sering digunakan untuk analisis komprehensif dari *dataset* pertambangan dan geologi yang besar. Misalnya, Caté *et al.* (2018), menggunakan algoritma klasifikasi *machine learning supervised* dengan *dataset* litogeokimia yang besar dari *logging* inti untuk membedakan unit batuan vulkanik dan tipe alterasi. Meskipun model tersebut dapat mengklasifikasikan unit vulkanik dengan akurat, akurasi untuk unit alterasi kurang memuaskan. Pembedaan tekstur dapat meningkatkan klasifikasi untuk unit alterasi tersebut, sementara penggunaan sifat batuan lainnya dapat membantu jika dua unit

batuan memiliki fitur tekstural yang identik. Dari berbagai penelitian ini, terlihat bahwa penggunaan berbagai algoritma *machine learning* dapat memberikan hasil yang signifikan dalam klasifikasi batuan dan mineral, namun pemilihan metode yang tepat serta karakteristik data yang digunakan memainkan peran penting dalam menentukan tingkat akurasinya.

Berdasarkan penjelasan yang telah disampaikan, penelitian ini akan dilakukan untuk mengklasifikasikan batuan beku berdasarkan data geokimia menggunakan algoritma *Random Forest Classifier*. Hasil penelitian ini diharapkan dapat membantu ahli geologi dalam melakukan penamaan batuan beku secara otomatis berdasarkan data geokimia.

2. METODE PENELITIAN

Secara garis besar, penelitian ini mencakup rangkaian langkah yang diambil selama pelaksanaan penelitian. Setiap tahap dalam proses penelitian dirancang dengan cermat untuk memastikan hasil yang valid dan dapat diandalkan. Proses penelitian ini dapat dilihat dalam Gambar 1 di bawah ini, yang menggambarkan alur kerja dari awal hingga akhir.



Gambar 1. Tahapan Penelitian

Langkah pertama melibatkan pengumpulan dan pemilihan *dataset* geokimia batuan beku yang relevan, diikuti oleh pengolahan dan pembersihan data untuk menghilangkan anomali dan memastikan konsistensi. Selanjutnya, *dataset* dibagi menjadi dua bagian: data latih (*training*) dan data uji (*testing*). Data latih digunakan untuk membangun model prediktif, sedangkan data uji digunakan untuk mengevaluasi kinerja model. Model prediktif kemudian dikembangkan menggunakan algoritma *Random Forest*, dioptimalkan, dan divalidasi menggunakan data latih. Langkah terakhir melibatkan pengujian model dengan data uji untuk mengevaluasi akurasi dan ketepatan prediksi.

2.1. Dataset Geokimia Batuan Beku

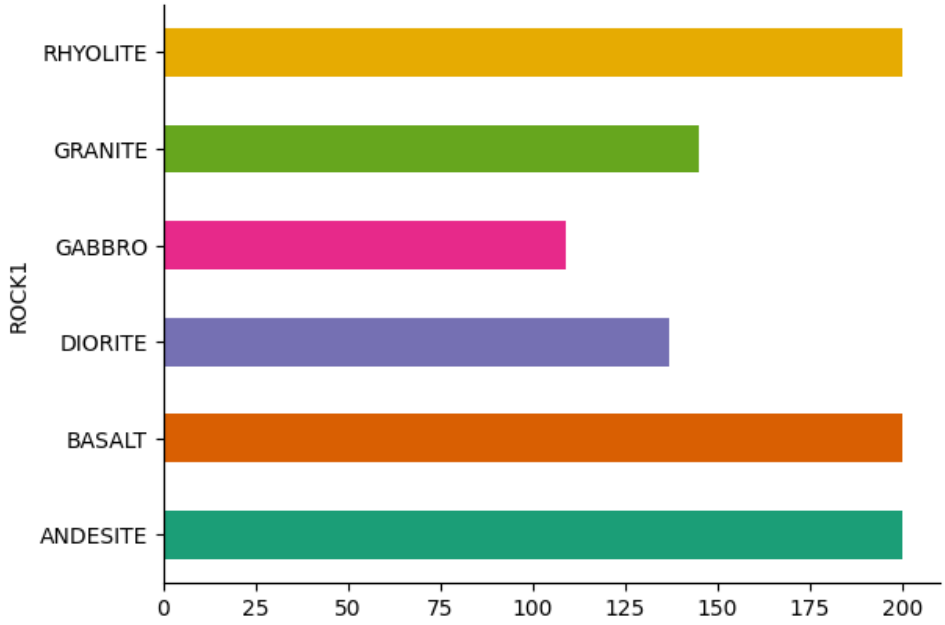
Dataset geokimia batuan beku yang digunakan dalam penelitian ini menggunakan *dataset* umum yang terbuka untuk publik berasal dari Cristian Cartagena Matos pada *website Kaggle* yang bisa diakses melalui tautan berikut: https://www.kaggle.com/datasets/cristianminas/geochemical-variations-in-igneous-rocks-mining. Penelitian ini menggunakan versi *dataset* yang telah diolah, terdiri dari 991 data yang dibagi menjadi dua bagian: 792 data (80%) untuk pelatihan (*training*) dan 199 data (20%) untuk pengujian

(*testing*). *Dataset* ini mencakup 11 variabel yang merupakan karakteristik *input*. Seperti yang ditunjukkan pada Tabel 1, *dataset* ini terdiri dari 12 variabel independen dan 1 variabel dependen. Variabel independen meliputi SIO2_WT%, TIO2_WT%, AL2O3_WT%, FEOT_WT%, CAO_WT%, MGO_WT%, MNO_WT%, K2O_WT%, NA2O_WT%, dan P2O5_WT%, sedangkan variabel dependen adalah target yang akan diprediksi, yaitu ROCK1.

Table 1. Deskripsi *Dataset*

Variabel	Deskripsi
ROCK1	Jenis Batuan
SIO2_WT%	Kandungan Silika / Silikon Dioksida pada Batuan dalam Berat Persen (wt.%)
TIO2_WT%	Kandungan Titanium Dioksida pada Batuan dalam Berat Persen (wt.%)
AL2O3_WT%	Kandungan Aluminium Oksida pada Batuan dalam Berat Persen (wt.%)
FEOT_WT%	Kandungan Besi Oksida pada batuan dalam Berat Persen (wt.%)
CAO_WT%	Kandungan Kalsium Oksida pada Batuan dalam Berat Persen (wt.%)
MGO_WT%	Kandungan Magnesium Oksida pada Batuan dalam Berat Persen (wt.%)
MNO_WT%	Kandungan Mangan(II) Oksida pada Batuan dalam Berat Persen (wt.%)
K2O_WT%	Kandungan Kalium Oksida pada Batuan dalam Berat Persen (wt.%)
NA2O_WT%	Kandungan Natrium Oksida pada Batuan dalam Berat Persen (wt.%)
P2O5_WT%	Kandungan Difosfor Pentaoksida pada Batuan dalam Berat Persen (wt.%)

Dataset terdiri dari enam kelas, dengan distribusi sebagai berikut: masing-masing 200 sampel untuk kelas *Andesite*, *Basalt*, dan *Rhyolite*; 145 sampel untuk kelas *Granite*; 137 sampel untuk kelas *Diorite*; dan 109 sampel untuk kelas *Gabbro*. Gambar 2 memperlihatkan distribusi kelas dalam dataset batuan beku ini.



Gambar 2. Distribusi kelas batuan beku

2.2. *Preprocessing Data*

Dataset yang didapat perlu menjalani tahap *preprocessing* sebelum dapat digunakan oleh sistem. Oleh karena itu, beberapa langkah *preprocessing* harus diterapkan untuk mengubah dan meningkatkan kualitas data. *Preprocessing* dilakukan untuk membersihkan data, menghilangkan *noise*, dan mengatasi nilai yang hilang sebelum langkah pemodelan dilakukan.

2.2.1. *Handling Missing Value*

Pada tahap awal *preprocessing*, data terlebih dahulu diproses untuk menangani nilai yang hilang (*missing value*). Nilai yang hilang dapat disebabkan oleh kesalahan *input* data atau data yang tidak tersedia, Karena algoritma *machine learning* tidak dapat bekerja dengan data yang memiliki nilai yang hilang. Masalah ini harus diselesaikan sebelum pemodelan dilakukan. Salah satu metode untuk mengatasi masalah *missing value* adalah imputasi. Imputasi adalah proses pengisian atau penggantian nilai-nilai yang hilang dalam suatu *dataset* dengan nilai-nilai yang dianggap masuk akal (*plausible values*) berdasarkan data yang tersedia dalam *dataset* tersebut (Myrtveit *et al*., 2001).

2.2.2. *Exploratory Data Analysis (EDA)*

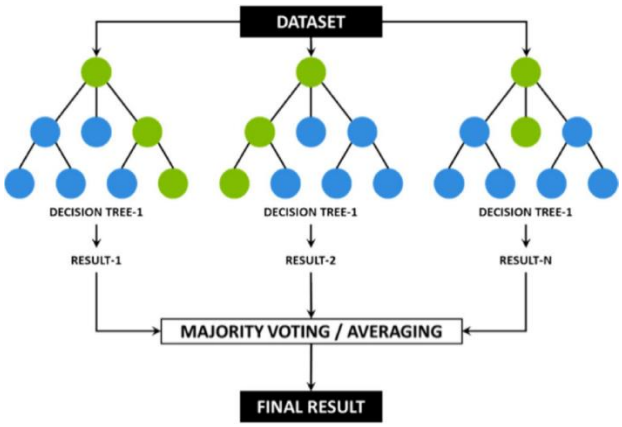
Exploratory Data Analysis (EDA) adalah langkah awal dalam penelitian yang bertujuan untuk mengidentifikasi pola, menemukan *outlier*, menguji hipotesis, dan memverifikasi asumsi. EDA sangat berguna untuk mendeteksi kesalahan sejak dini, karena memungkinkan pengguna menemukan anomali dalam data, memahami hubungan antar data, dan mengekstrak faktor-faktor penting. Proses EDA ini sangat bermanfaat dalam analisis statistik.

2.2.3. *Pembagian Data*

Tahap berikutnya adalah membagi data menjadi data *training* dan data *testing*. Pada tahap ini, data *training* digunakan untuk membangun model atau pola, sedangkan data *testing* digunakan untuk mengevaluasi kinerja model tersebut.

2.3. *Implementasi Random Forest Classifier*

Random Forest merupakan sebuah algoritma dalam *supervised learning* yang dapat diterapkan untuk melakukan klasifikasi atau regresi. Algoritma ini terkenal karena fleksibilitasnya dan kemudahan penggunaannya. *Random Forest* (RF) terdiri dari sejumlah *Decision Tree*. Semakin banyak *Decision Tree* yang terlibat, semakin kuat pula kinerja algoritma *Random Forest* tersebut (Gambar 2).



Gambar 2. Algoritma *Random Forest* (Alhams *et al*., 2024)

Algoritma *Random Forest* memanfaatkan metode rata-rata untuk meningkatkan akurasi prediksi dan mengendalikan *overfitting*. Pengaturan ukuran sub-sampel dapat dikontrol dengan parameter *max samples* ketika *bootstrap=True* (*default*). Jika tidak, keseluruhan *dataset* digunakan untuk membangun setiap pohon (Buitinck *et al*., 2011). Saat menggunakan *Random Forest* untuk melakukan klasifikasi data, formula Indeks *Gini* seperti yang bisa dilihat dalam persamaan (1), digunakan untuk menentukan cara *node* pada sebuah cabang pohon keputusan diatur. Persamaan ini memperhitungkan kelas dan probabilitas untuk menentukan *Gini* dari setiap cabang di sebuah simpul, sehingga memutuskan cabang mana yang lebih mungkin terjadi.

Struktur pohon keputusan ini terbagi menjadi tiga bagian utama: *root node*, *internal node*, dan *leaf node*. Penentuan simpul atau akar dari pohon keputusan dapat menggunakan nilai *entropy*, seperti yang dinyatakan dalam persamaan (1), serta nilai *information gain*, sebagaimana dijelaskan dalam persamaan (2), atau menggunakan indeks *Gini*, sebagaimana dijelaskan dalam persamaan (3), dan *Gini split*, sebagaimana dijelaskan dalam persamaan (4). Proses kerja *Random Forest* dimulai dengan menentukan jumlah pohon yang akan dibuat, yaitu n. Kemudian, dilakukan proses *bagging* dengan mengambil sampel fitur dan baris data untuk membangun model menggunakan *decision tree*, sehingga tercipta sejumlah pohon yang telah ditentukan. Selanjutnya *entropy*, *information gain*, atau *gini split index* digunakan untuk membangun pohon dan mengevaluasi hasil prediksi mayoritas (Ferrer dan Aragón, 2023).

$$Entropy\ (S) = \sum_{i=1}^{n} -P_i \log_2 P_i \quad (1)$$

$$Gain\ (S, J) = Entropy(S) - \sum_{i=1}^{n} \frac{|S_i|}{|S|} \times Entropy(S_i) \quad (2)$$

$$Gain\ (S) = 1 - \sum_{i=1}^{n} (P_i{}^2) \quad (3)$$

$$Gini_{split} = \sum_{i=1}^{n} \frac{|S_i|}{|S|} \times Gini(S_i) \quad (4)$$

2.4. Evaluasi Hasil Model

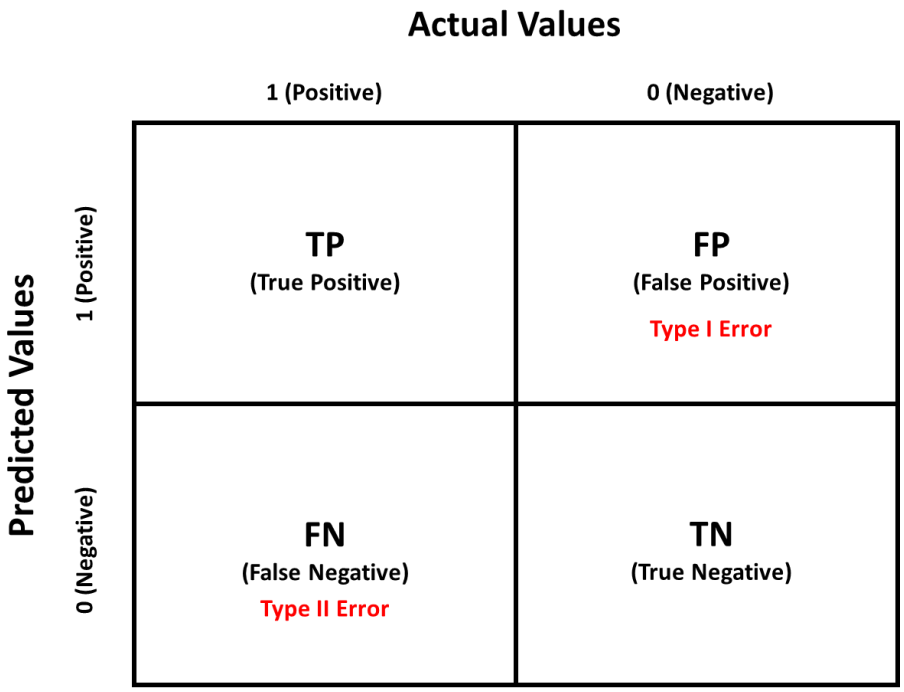
Memilih formula kinerja yang sesuai untuk mengevaluasi algoritma adalah langkah kritis, karena klasifikasi yang dilatih pada *dataset* yang tidak seimbang dapat memberikan akurasi yang tinggi, tetapi sebenarnya cenderung memihak kepada kelas mayoritas. Penggunaan formula kinerja yang tepat akan membantu dalam menilai kemampuan adaptasi algoritma secara efisien. Tujuan utamanya adalah untuk mencapai sebanyak mungkin *True Positive* (TP) dan *True Negative* (TN) sambil juga mengurangi *False Negative* sebanyak mungkin. Akurasi (*Accuracy*) mencerminkan kinerja keseluruhan pengklasifikasi. Metrik kinerja lainnya adalah *Recall/Sensitivity*, yang mengukur akurasi kelas positif, dan *Spesificity*, yang mengukur akurasi kelas negatif. *Sensitivity* mengevaluasi efektivitas *classifier* pada kelas positif/mayoritas, sementara *Spesificity* mengevaluasi efektivitas *classifier* pada kelas negatif/mayoritas. *Precision*, juga merupakan metrik kinerja penting, adalah ukuran ketepatan model. Presisi yang tinggi dari sebuah *classifier* menunjukkan bahwa *classifier* tersebut baik (Turlapati dan Prusty, 2020).

Pada tahap ini, akan disusun sebuah *Confusion matrix* untuk menilai kinerja model *machine learning*. *Confusion matrix* merupakan tabel yang digunakan untuk mengevaluasi performa model klasifikasi *machine learning* dengan menyajikan hasil prediksi model pada suatu *dataset* (Gambar 3). Ini memungkinkan kita untuk menilai seberapa akurat atau tidak akurat model dalam mengklasifikasikan data.

Dalam menggunakan *Confusion Matrix* untuk menilai kinerja metode klasifikasi, ada empat istilah yang mewakili hasil dari proses klasifikasi. Keempat representasi hasil proses klasifikasi tersebut ditampilkan dalam Tabel 2.

Table 2. Representasi Hasil Proses Klasifikasi *Confusion Matrix* (Santra & Christy, 2012)

True Positive **(TP)**	Total data positif yang terklasifikasi dengan benar oleh sistem
True Negative **(TN)**	Total data negatif yang terklasifikasi dengan benar oleh sistem
False Negative **(FN)**	Total data positif akan tetapi terklasifikasi sebagai data negatif oleh sistem
False Positive **(FP)**	Total data negatif akan tetapi terklasifikasi sebagai data positif oleh sistem



Gambar 3. *Confusion Matrix* (Siregar *et al.*, 2023)

Evaluasi kinerja model dalam penelitian ini melibatkan metrik seperti *Accuracy*, Ketepatan (*Precision*), *Recall*, dan *F1-score*. Akurasi mengindikasikan seberapa baik model dapat mengklasifikasikan data secara tepat. Formula untuk menghitung akurasi bisa dilihat pada persamaan (5) sebagai berikut:

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \times 100\% \quad (5)$$

Precision digunakan untuk menilai sejauh mana prediksi positif model adalah benar. Formula untuk menghitung *precision* adalah sebagai berikut (6):

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

Recall digunakan untuk menilai sejauh mana model mampu mengidentifikasi dan mengklasifikasikan semua kasus positif dengan benar. Rumus untuk menghitung *recall* adalah sebagai berikut (7):

$$Recall = \frac{TP}{TP+FN} \times 100\% \quad (7)$$

F1-score digunakan untuk menunjukkan perbandingan yang dibobot antara *precision* dan *recall* secara rata-rata. Formula untuk menghitung *F1-score* adalah sebagai berikut (8):

$$F1 - Score = \frac{2*Precision*Recall}{Precision+Recall} \quad (8)$$

3. HASIL DAN PEMBAHASAN

3.1. *Dataset*

Dataset geokimia batuan beku yang digunakan berasal dari *Kaggle*. Pada Gambar 4 memberikan informasi umum tentang *dataset* ini, yang terdiri dari 11 kolom (variabel). Terdapat 10 variabel dengan tipe data *float*, yaitu SIO2_WT%, TIO2_WT%, AL2O3_WT%, FEOT_WT%, CAO_WT%, MGO_WT%, MNO_WT%, K2O_WT%, NA2O_WT%, dan P2O5_WT%. Satu variabel lainnya, ROCK1, memiliki tipe data *object*. Total jumlah sampel dalam *dataset* ini adalah 991, namun beberapa variabel memiliki nilai yang hilang sehingga jumlah sampelnya kurang dari 991. Variabel tersebut adalah FEOT_WT% dengan 943 sampel, CAO_WT% dengan 979 sampel, MNO_WT% dengan 974 sampel, K2O_WT% dengan 984 sampel, NA2O_WT% dengan 990 sampel, dan P2O5_WT% dengan 942 sampel. Oleh karena itu, perlu dilakukan *preprocessing* terhadap nilai yang hilang (*missing values*) dalam *dataset* ini.

	ROCK1	SIO2_WT%	TIO2_WT%	AL2O3_WT%	FEOT_WT%	CAO_WT%	MGO_WT%	MNO_WT%	K2O_WT%	NA2O_WT%	P2O5_WT%
0	ANDESITE	58.60	1.61	12.83	11.53	6.52	4.87	0.20	0.16	2.51	0.16
1	ANDESITE	57.30	1.82	12.33	12.73	5.68	4.10	0.20	0.20	3.73	0.19
2	ANDESITE	55.00	1.32	13.80	11.82	7.66	6.05	0.20	0.22	3.10	0.13
3	ANDESITE	57.50	1.18	12.80	11.82	10.70	6.11	0.24	0.11	1.48	0.13
4	ANDESITE	57.50	1.64	12.89	11.84	7.87	3.41	0.22	0.49	1.89	0.17
...	...	...	...	...	...	...	...	...	...	...	...
986	DIORITE	63.70	0.65	16.67	4.75	0.09	2.98	5.70	3.83	1.45	0.17
987	DIORITE	56.77	1.28	17.32	8.79	0.17	3.26	6.63	4.47	1.07	0.25
988	DIORITE	52.67	1.01	17.87	8.45	0.16	6.15	10.27	2.67	0.56	0.18
989	DIORITE	62.66	0.82	16.16	5.73	0.10	3.20	5.28	3.23	2.60	0.21
990	DIORITE	64.36	0.62	17.22	4.59	0.08	2.20	5.21	4.07	1.46	0.19

991 rows × 11 columns

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 991 entries, 0 to 990
Data columns (total 11 columns):
 #   Column     Non-Null Count  Dtype
---  ------     --------------  -----
 0   ROCK1      991 non-null    object
 1   SIO2_WT%   991 non-null    float64
 2   TIO2_WT%   991 non-null    float64
 3   AL2O3_WT%  991 non-null    float64
 4   FEOT_WT%   943 non-null    float64
 5   CAO_WT%    979 non-null    float64
 6   MGO_WT%    991 non-null    float64
 7   MNO_WT%    974 non-null    float64
 8   K2O_WT%    984 non-null    float64
 9   NA2O_WT%   990 non-null    float64
 10  P2O5_WT%   942 non-null    float64
dtypes: float64(10), object(1)
memory usage: 85.3+ KB
```

Gambar 4. Informasi *Dataset*

3.2. *Preprocessing*

3.2.1. *Handling Missing Values*

Pada tahap ini melakukan *handling missing values* untuk mengidentifikasi dan penanganan terhadap kesalahan, inkonsistensi, dan ketidakakuratan yang ada dalam *dataset* yang digunakan. Dalam penelitian ini proses *data cleaning* yang digunakan yaitu dengan cara mengisi nilai yang hilang (*missing value*) pada *dataset* dengan nilai 0. Dapat dilihat pada tabel 3 terdapat *missing value* pada variabel FEOT_WT% dengan 48 sampel, CAO_WT% dengan 12 sampel, MNO_WT% dengan 17 sampel, K2O_WT% dengan 7 sampel, NA2O_WT% dengan 1 sampel, dan P2O5_WT% dengan 49 sampel. Berdasarkan hal itu, untuk mengatasi nilai yang hilang (*missing value*) yaitu dengan cara mengisi baris atau kolom yang mengandung nilai yang hilang tersebut. Setelah melakukan *handling missing values* jumlah *record* data untuk masing-masing variabel menjadi 991 baris.

```
ROCK1          0
SIO2_WT%       0
TIO2_WT%       0
AL2O3_WT%      0
FEOT_WT%      48
CAO_WT%       12
MGO_WT%        0
MNO_WT%       17
K2O_WT%        7
NA2O_WT%       1
P2O5_WT%      49
dtype: int64
```

```
ROCK1         0
SIO2_WT%      0
TIO2_WT%      0
AL2O3_WT%     0
FEOT_WT%      0
CAO_WT%       0
MGO_WT%       0
MNO_WT%       0
K2O_WT%       0
NA2O_WT%      0
P2O5_WT%      0
dtype: int64
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 991 entries, 0 to 990
Data columns (total 11 columns):
 #   Column     Non-Null Count  Dtype
---  ------     --------------  -----
 0   ROCK1      991 non-null    object
 1   SIO2_WT%   991 non-null    float64
 2   TIO2_WT%   991 non-null    float64
 3   AL2O3_WT%  991 non-null    float64
 4   FEOT_WT%   991 non-null    float64
 5   CAO_WT%    991 non-null    float64
 6   MGO_WT%    991 non-null    float64
 7   MNO_WT%    991 non-null    float64
 8   K2O_WT%    991 non-null    float64
 9   NA2O_WT%   991 non-null    float64
 10  P2O5_WT%   991 non-null    float64
dtypes: float64(10), object(1)
memory usage: 85.3+ KB
```

Gambar 5. *Hasil Handling Missing Value*. Kiri : *Missing value* tiap variabel. Tengah : Setelah *handling missing value*. Kanan : Jumlah total sampel tiap variabel setelah dilakukan *handing missing values*.

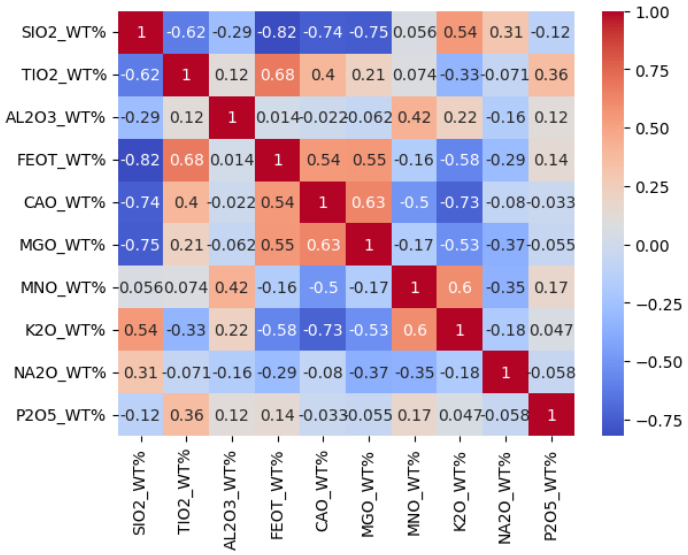
3.2.2. Exploratory Data Analysis

Exploratory Data Analysis (EDA) merupakan langkah penting sebelum melakukan pemodelan data. Melalui proses EDA, dapat dipahami secara utuh data yang ada. Pada Gambar 6 menyajikan statistik deskriptif untuk variabel-variabel geokimia dalam dataset batuan beku. Semua variabel memiliki jumlah sampel yang sama, yaitu 991. SIO2_WT% memiliki rata-rata tertinggi (61.244805), menunjukkan bahwa komponen ini paling dominan dalam data batuan beku tersebut, sedangkan P2O5_WT% memiliki rata-rata terendah (0.123751), menunjukkan bahwa komponen ini paling sedikit terdapat dalam dataset batuan beku. SIO2_WT% juga memiliki standar deviasi tertinggi (10.727290), menunjukkan variasi yang signifikan dalam kandungan SIO2_WT% di antara sampel, sementara P2O5_WT% memiliki standar deviasi terendah (0.163556), menunjukkan bahwa kandungannya cukup konsisten di antara sampel. Nilai minimum untuk semua variabel adalah 0.000000 kecuali SIO2_WT% dan AL2O3_WT%. Nilai maksimum berkisar dari 3.200000 untuk TIO2_WT% hingga 90.100000 untuk SIO2_WT%. Kuartil pertama (25%), median (50%), dan kuartil ketiga (75%) memberikan informasi tentang distribusi data; misalnya, untuk SIO2_WT%, 50% sampel berada di bawah nilai 59.560000, dan untuk TIO2_WT%, 50% sampel berada di bawah nilai 0.710000. Variabel lain juga menunjukkan variasi dalam distribusi konsentrasi mereka di batuan beku. Secara keseluruhan, Gambar 6 ini menunjukkan bahwa terdapat variasi yang cukup besar dalam kandungan kimia dari berbagai unsur di batuan beku, dengan beberapa unsur memiliki variasi yang lebih tinggi dibandingkan yang lain. SIO2_WT% adalah komponen yang paling dominan, sementara P2O5_WT% adalah yang paling sedikit ditemukan.

	SIO2_WT%	TIO2_WT%	AL2O3_WT%	FEOT_WT%	CAO_WT%	MGO_WT%	MNO_WT%	K2O_WT%	NA2O_WT%	P2O5_WT%
count	991.000000	991.000000	991.000000	991.000000	991.000000	991.000000	991.000000	991.000000	991.000000	991.000000
mean	61.244805	0.780568	14.546364	6.996771	4.515123	3.972880	1.302060	1.830894	2.643992	0.123751
std	10.727290	0.525606	2.538211	5.182870	4.359360	3.692658	2.231128	1.713481	1.466493	0.163556
min	20.640000	0.000000	3.230000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	52.200000	0.337000	13.015000	2.775000	0.120000	1.000000	0.110000	0.270000	1.425000	0.040000
50%	59.560000	0.710000	14.570000	6.110000	3.110000	3.090000	0.200000	1.030000	2.700000	0.090000
75%	71.895000	1.035000	16.360000	10.680000	8.500000	5.805000	1.570000	3.625000	3.750000	0.180000
max	90.100000	3.200000	23.600000	39.430000	18.990000	25.900000	10.270000	8.250000	7.250000	4.030000

Gambar 6. Statistik Dataset

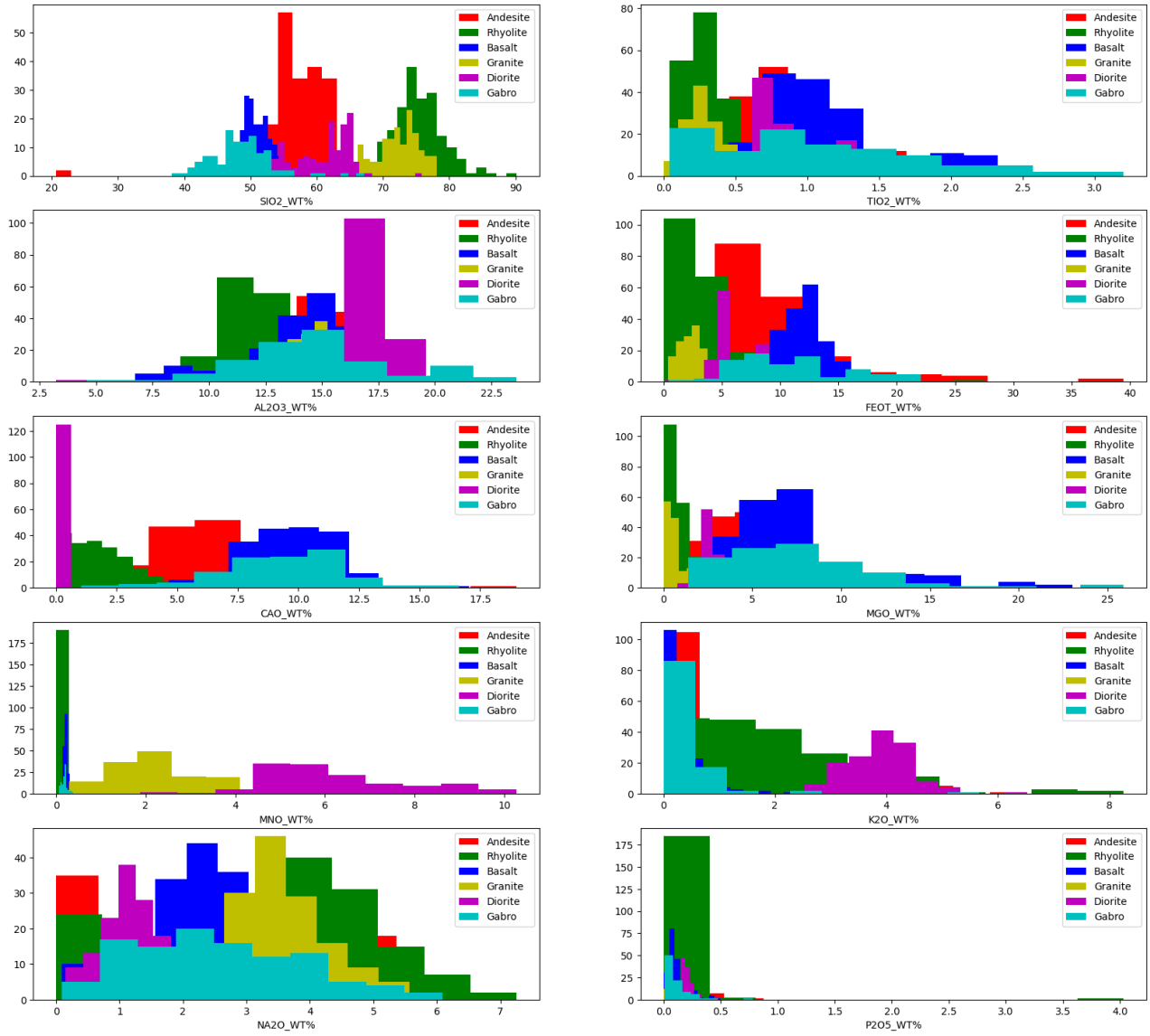
Visualisasi merupakan langkah terpenting dalam EDA yang menunjukkan keterkaitan antar variabel independen dengan target. Pada Gambar 7 menunjukkan korelasi antara variabel-variabel geokimia dalam *dataset* batuan beku. SIO2_WT% memiliki korelasi negatif yang kuat dengan beberapa variabel lain, seperti TIO2_WT% (-0.62), FEOT_WT% (-0.82), CAO_WT% (-0.74), dan MGO_WT% (-0.75), menunjukkan bahwa ketika kandungan SIO2_WT% meningkat, kandungan variabel-variabel ini cenderung menurun. Sebaliknya, FEOT_WT% memiliki korelasi positif yang kuat dengan TIO2_WT% (0.68), CAO_WT% (0.54) dan MGO_WT% (0.55), serta CAO_WT% juga memiliki korelasi positif yang kuat dengan MGO_WT% (0.63), menunjukkan bahwa ketika kandungan FEOT_WT% meningkat, kandungan TIO2_WT%, CAO_WT% dan MGO_WT% juga cenderung meningkat. Beberapa variabel menunjukkan korelasi yang sangat rendah atau hampir nol, seperti SIO2_WT% dengan AL2O3_WT% (-0.29) dan SIO2_WT% dengan P2O5_WT% (-0.12), menunjukkan bahwa perubahan dalam kandungan SIO2_WT% tidak banyak berpengaruh pada kandungan AL2O3_WT% dan P2O5_WT%. Terdapat berbagai tingkat korelasi antara variabel lain, seperti AL2O3_WT% dengan TIO2_WT% (0.12) dan AL2O3_WT% dengan P2O5_WT% (0.14), yang menunjukkan hubungan yang lebih lemah namun masih ada. Secara keseluruhan, *heatmap* ini memberikan gambaran jelas tentang bagaimana variabel-variabel geokimia ini saling berkorelasi satu sama lain dalam dataset batuan beku, dengan beberapa variabel menunjukkan hubungan yang kuat baik secara positif maupun negatif.



Gambar 7. *Heatmap* korelasi antar variabel

Histogram pada Gambar 8 menunjukkan distribusi konsentrasi berbagai elemen geokimia dalam batuan beku, dikelompokkan berdasarkan jenis batuan (*Andesite*, *Rhyolite*, *Basalt*, *Granite*, *Diorite*, dan *Gabbro*). Distribusi SIO2_WT% menunjukkan *Andesite* dan *Basalt* memiliki puncak di sekitar 50-55%, sementara *Rhyolite* dan *Granite* berada di sekitar 70-75%. TIO2_WT% pada semua jenis batuan memiliki puncak di nilai rendah (0-1%), dengan *Basalt* sedikit bervariasi ke nilai yang lebih tinggi. AL2O3_WT% menunjukkan puncak pada *Gabbro* di sekitar 15-17%, sedangkan *Andesite*, *Basalt*, dan *Granite* memiliki puncak di sekitar 14-16%. Distribusi FEOT_WT% lebih lebar pada *Gabbro* dan *Basalt* dengan puncak di sekitar 5-10%, sedangkan *Rhyolite* dan *Granite* memiliki nilai yang lebih rendah. CAO_WT% pada Gabbro memiliki puncak di sekitar 10-12%, sementara *Andesite*, *Basalt*, dan *Granite* berada di sekitar 4-8%. MGO_WT% menunjukkan *Gabbro* dengan puncak di sekitar 5-10%, dan *Andesite* serta *Basalt* di sekitar 2-4%. MNO_WT% menunjukkan semua batuan dengan puncak di nilai rendah (0-2%). K2O_WT% memiliki puncak pada *Rhyolite* dan *Granite* di sekitar 2-4%, dan *Andesite* serta *Basalt* di sekitar 0-1%. NA2O_WT% pada *Andesite* dan *Granite* berada di sekitar 3-4%, sedangkan *Basalt* dan *Gabbro* menunjukkan distribusi

lebih luas di sekitar 2-4%. P2O5_WT% menunjukkan semua batuan dengan distribusi di nilai rendah (0-0.5%), dengan sedikit variasi lebih tinggi pada *Basalt* dan *Gabbro*. Secara keseluruhan, histogram ini mengindikasikan bahwa setiap jenis batuan beku memiliki distribusi elemen geokimia yang berbeda, *Andesite* dan *Basalt* cenderung memiliki konsentrasi yang lebih rendah untuk sebagian besar elemen, sementara *Rhyolite* dan *Granite* cenderung memiliki konsentrasi yang lebih tinggi, terutama untuk SIO2_WT% dan K2O_WT%. *Gabbro* menunjukkan distribusi yang lebih lebar dan bervariasi untuk beberapa elemen seperti CAO_WT% dan FEOT_WT%.



Gambar 8. Distribusi konsentrasi berbagai elemen geokimia dalam batuan beku

3.2.3. Pembagian Data

Sebelum melakukan penerapan algoritma *Random Forest Classifier*, perlu melakukan pembagian data atau *split* data menjadi dua bagian, yaitu data latih dan data uji. Pada penelitian ini akan menggunakan *split* data dengan proporsi 80:20 digunakan untuk membagi data menjadi data latih sebesar 80% dan data uji sebesar 20%. Pada kasus ini, jumlah data yang akan diolah ada sebanyak 991 data, sehingga 80% diantaranya yaitu

sebanyak 792 data sebagai data latih (Gambar 9), dan 20% sisanya yaitu sebanyak 199 data sebagai data uji (Gambar 10).

	SIO2_WT%	TIO2_WT%	AL2O3_WT%	FEOT_WT%	CAO_WT%	MGO_WT%	MNO_WT%	K2O_WT%	NA2O_WT%	P2O5_WT%
948	59.55	0.94	16.98	7.33	0.13	3.60	6.91	3.27	1.11	0.17
774	73.56	0.26	14.47	1.70	0.02	0.45	2.17	4.05	3.26	0.06
43	58.10	1.00	17.90	3.87	5.60	1.10	0.20	2.00	4.90	0.40
976	65.98	0.61	17.43	4.19	0.08	1.71	5.57	3.32	0.95	0.16
805	75.32	0.19	13.05	1.70	0.03	0.25	1.21	3.52	4.70	0.03
...	...	...	...	...	...	...	...	...	...	...
890	64.05	0.62	16.87	4.80	0.09	2.24	6.47	3.20	1.43	0.23
902	60.03	0.92	17.55	6.01	0.10	3.66	6.04	4.34	1.10	0.24
661	46.94	1.21	16.63	14.15	11.43	5.77	0.23	0.28	1.70	0.08
607	51.60	1.36	15.20	12.78	8.70	5.10	0.27	0.30	3.20	0.12
803	75.21	0.15	13.55	1.32	0.03	0.90	1.33	3.31	4.17	0.03

792 rows × 10 columns

Gambar 9. Data *Training*

	SIO2_WT%	TIO2_WT%	AL2O3_WT%	FEOT_WT%	CAO_WT%	MGO_WT%	MNO_WT%	K2O_WT%	NA2O_WT%	P2O5_WT%
419	73.62	0.12	10.32	3.65	2.80	0.43	0.00	0.91	4.20	0.00
655	53.60	0.89	16.36	10.28	9.70	6.09	0.18	0.01	1.67	0.05
57	59.34	0.99	14.80	7.36	5.95	4.32	0.19	0.13	4.29	0.19
301	45.61	2.08	13.31	18.80	9.92	2.91	0.34	0.87	3.67	0.29
851	67.13	0.54	15.81	3.83	0.09	1.44	3.58	4.14	3.34	0.11
...	...	...	...	...	...	...	...	...	...	...
353	80.77	0.11	11.15	1.67	1.35	0.74	0.11	3.62	0.26	0.01
977	58.41	0.61	16.64	6.52	0.17	4.97	7.36	3.34	1.85	0.12
576	48.42	1.01	14.54	13.82	9.75	7.42	0.25	0.61	2.58	0.07
63	55.93	0.67	14.14	9.01	9.54	6.55	0.17	0.58	2.35	0.07
50	68.70	0.50	15.80	6.30	1.00	1.60	0.10	2.00	3.20	0.10

199 rows × 10 columns

Gambar 10. Data *Testing*

3.3. Implementasi Random Forest Classifier

Klasifikasi dilakukan dengan menggunakan tuning *hyperparameter* pada *Random Forest Classifier*. Pada penelitian ini, metode *Grid Search* digunakan untuk melakukan tuning *hyperparameter*. Tujuan dari tuning *hyperparameter* ini adalah untuk mengurangi kemungkinan model menjadi terlalu *overfit* saat mengklasifikasi batuan.

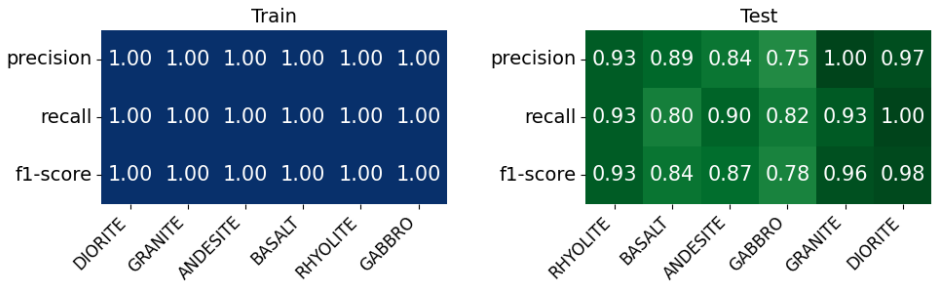
Table 3. Hyperparameter pada Random Forest Classifier

Hyperparameter	***Grid Search Values***	**Nilai Parameter Terbaik**
n_estimators	[100, 150, 200]	200
max_depth	[20, 50, 80]	20
max_features	[0.3, 0.6, 0.8]	0.3
min_samples_leaf	[1, 5, 10]	1

Tabel 3 menunjukkan parameter dan nilai parameter optimal yang dapat meningkatkan kinerja algoritma dalam eksperimen yang dilakukan. Model dengan akurasi terbaik adalah *Random Forest Classifier* dengan parameter *n_estimators* sebesar 200 pohon keputusan, *max_depth* sebesar 20, *max_features* sebesar 0.3, dan *min_samples_leaf* sebesar 1. Model terbaik ini mencapai akurasi sebesar 89.44%.

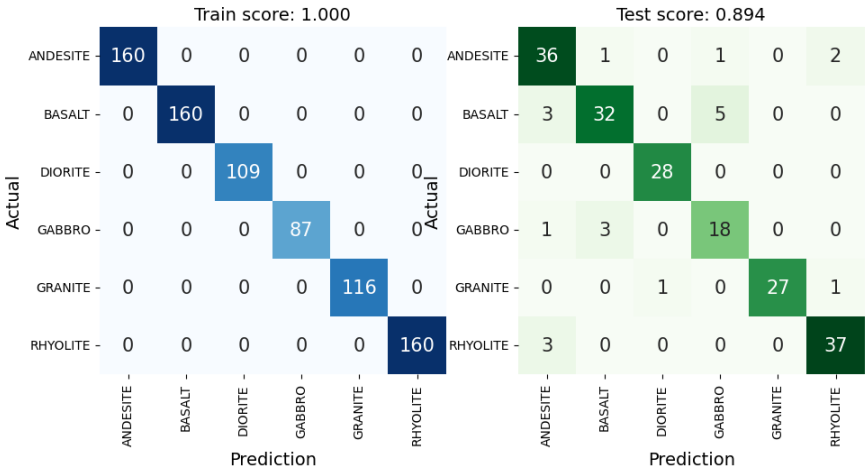
3.4. Evaluasi Hasil Model

Evaluasi model merupakan suatu proses untuk mengukur seberapa baik kinerja model yang telah dibangun. Hal ini penting untuk memahami sejauh mana model dapat menggeneralisasi pada data yang belum pernah dilihat sebelumnya. Pada penelitian ini didapatkan akurasi sebesar 0.894 atau 89.4%.



Gambar 11. *Classification Report*

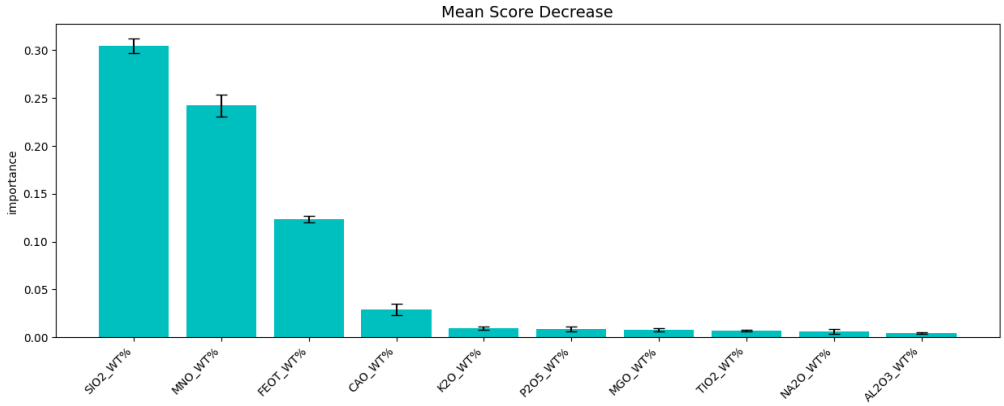
Pada Gambar 11 menampilkan metrik *classification report* (*precision*, *recall*, dan *f1-score*) dari model klasifikasi batuan pada data pelatihan (*Train*) dan data pengujian (*Test*). Pada data pelatihan, *precision*, *recall*, dan *f1-score* untuk semua jenis batuan (*Diorite*, *Granite*, *Andesite*, *Basalt*, *Rhyolite*, dan *Gabbro*) adalah 1.00, menunjukkan bahwa model sepenuhnya akurat dalam mengklasifikasi semua jenis batuan pada data pelatihan tanpa kesalahan. Namun, pada data pengujian, nilai *precision* berkisar antara 0.75 (untuk *Gabbro*) hingga 1.00 (untuk *Granite* dan *Diorite*), nilai *recall* berkisar antara 0.80 (untuk *Basalt*) hingga 1.00 (untuk *Diorite*), dan nilai *f1-score* berkisar antara 0.78 (untuk *Gabbro*) hingga 0.98 (untuk *Diorite*). Metrik *classification report* ini menunjukkan bahwa performa model pada data pengujian sedikit lebih rendah dibandingkan pada data pelatihan, dengan beberapa jenis batuan seperti *Gabbro* menunjukkan nilai yang lebih rendah.



Gambar 12. *Confusion Matrix model*

Terlihat pada Gambar 12 menampilkan *confusion matrix* untuk model klasifikasi batuan pada data pelatihan (*Train*) dan data pengujian (*Test*). Pada data pelatihan, model berhasil mengklasifikasi semua sampel dengan

benar untuk setiap jenis batuan yang ditunjukkan dengan nilai sempurna pada diagonal utama dan skor pelatihan 1.00. Pada data pengujian, terdapat beberapa kesalahan klasifikasi. Sebagai contoh, dari 40 sampel *Andesite*, 36 diklasifikasikan dengan benar, sementara 1 sampel diklasifikasikan sebagai *Basalt*, 1 sebagai *Diorite*, dan 2 sebagai *Rhyolite*. Kesalahan lainnya termasuk 5 sampel *Basalt* yang diklasifikasikan sebagai *Granite*, 3 sampel *Gabbro* sebagai *Basalt*, dan sebagainya. Skor pengujian adalah 0.894, menunjukkan bahwa model masih cukup akurat pada data pengujian.



Gambar 13. *Features Importance*

Pada Gambar 13 menunjukkan diagram batang yang mengilustrasikan penurunan skor rata-rata (*Mean Score Decrease*) untuk setiap variabel (atau fitur) dalam model klasifikasi batuan, yang merepresentasikan pentingnya masing-masing variabel dalam model. Variabel paling penting dalam model klasifikasi batuan ini adalah variabel SIO2_WT%, dengan penurunan skor rata-rata terbesar sekitar 0.30, diikuti oleh MNO_WT% dan FEOT_WT% dengan penurunan skor rata-rata masing-masing sekitar 0.25 dan 0.13. CAO_WT% memiliki penurunan skor rata-rata sekitar 0.03, menunjukkan tingkat pentingnya yang moderat. Variabel lain seperti K2O_WT%, P2O5_WT%, MGO_WT%, TIO2_WT%, NA2O_WT%, dan AL2O3_WT% memiliki penurunan skor rata-rata yang jauh lebih kecil, menunjukkan kontribusi yang lebih rendah terhadap model.

KESIMPULAN

a) Model menunjukkan performa yang cukup baik dengan akurasi 89.4% pada data *test*.
b) Pada data *training*, *precision*, *recall*, dan *f1-score* untuk semua jenis batuan adalah 1.00, menunjukkan akurasi sempurna. sedangkan pada data *test*, nilai *precision* berkisar antara 0.75 hingga 1.00, *recall* antara 0.80 hingga 1.00, dan *f1-score* antara 0.78 hingga 0.98.
c) Variabel paling penting dalam model klasifikasi batuan ini adalah SIO2_WT%, dengan penurunan skor rata-rata terbesar sekitar 0.30, diikuti oleh MNO_WT% (0.25) dan FEOT_WT% (0.13). Variabel lain seperti CAO_WT%, K2O_WT%, P2O5_WT%, MGO_WT%, TIO2_WT%, NA2O_WT%, dan AL2O3_WT% memiliki penurunan skor rata-rata yang lebih kecil, menunjukkan kontribusi yang lebih rendah.

DAFTAR PUSTAKA

Alférez, G. H., Serrano, S. H., Ardila, A. M. M., & Clausen, B. L. (2021). Automatic Classification of Plutonic Rocks with Machine Learning Applied to Extracted Shades and Colors on iOS Devices. *Proceedings of the Future Technologies Conference (FTC)*, 72–88.

Alhams, A., Abdelhadi, A., Badri, Y., Sassi, S., & Renno, J. (2024). Enhanced Bearing Fault Diagnosis Through Trees Ensemble Method and Feature Importance Analysis. *Journal of Vibration Engineering and Technologies*,

0123456789. https://doi.org/10.1007/s42417-024-01405-0

Bamford, T., Esmaeili, K., & Schoellig, A. P. (2021). A deep learning approach for rock fragmentation analysis. *International Journal of Rock Mechanics and Mining Sciences*, *145*. https://doi.org/10.1016/j.ijrmms.2021.104839

Buitinck, L., Louppe, G., Blondel, M., Fabien, P., Mueller, A., Olivier, G., Niculae, V., Prettenhofer, P., Gramfort, A., Grobler, J., Layton, R., VanderPlas, J., Joly, A., & Varoquaux, G. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830. http://scikit-learn.sourceforge.net.

Caté, A., Schetselaar, E., Mercier-Langevin, P., & Ross, P. S. (2018). Classification of lithostratigraphic and alteration units from drillhole lithogeochemical data using machine learning: A case study from the Lalor volcanogenic massive sulphide deposit, Snow Lake, Manitoba, Canada. *Journal of Geochemical Exploration*, *188*, 216–228. https://doi.org/10.1016/j.gexplo.2018.01.019

Ferrer, C. A., & Aragón, E. (2023). Note on "A Comprehensive Analysis of Synthetic Minority Oversampling Technique (SMOTE) for handling class imbalance." *Information Sciences*, *630*, 322–324. https://doi.org/10.1016/j.ins.2022.10.005

Houshmand, N., GoodFellow, S., Esmaeili, K., & Ordóñez Calderón, J. C. (2022). Rock type classification based on petrophysical, geochemical, and core imaging data using machine and deep learning techniques. *Applied Computing and Geosciences*, *16*. https://doi.org/10.1016/j.acags.2022.100104

Kantu, R. R. (2022). Geokimia Batuan Kubah Lava Daerah Bulu Dua Kecamatan Tanete Riaja Provinsi Sulawesi Selatan [Universitas Hasanuddin]. In *Doctoral dissertation, Universitas Hasanuddin* (Issue 8.5.2017). www.aging-us.com

Li, X., & Li, H. (2013). A new method of identification of complex lithologies and reservoirs: task-driven data mining. *J. Petrol. Sci. Eng*, *109*, 241–249. https://doi.org/10.1016/%0Aj.petrol.2013.08.049

Mangalathu, S., & Burton, H. V. (2019). Deep learning-based classification of earthquake-impacted buildings using textual damage descriptions. *International Journal of Disaster Risk Reduction*, *36*. https://doi.org/10.1016/j.ijdrr.2019.101111

Myrtveit, I., Stensrud, E., & Olsson, U. H. (2001). Analyzing data sets with missing data: An empirical evaluation of imputation methods and likelihood-based methods. *IEEE Transactions on Software Engineering*, *27*(11), 999–1013. https://doi.org/10.1109/32.965340

Niu, G., He, X., Xu, H., & Dai, S. (2024). Development of Rock Classification Systems: A Comprehensive Review with Emphasis on Artificial Intelligence Techniques. *Eng*, *5*(1), 217–245. https://doi.org/10.3390/eng5010012

Pane, S. A., & Sihombing, F. M. H. (2020). *No TitleKlasifikasi dan klasterisasi mineral batuan di lapangan x berdasarkan data spektral (SWIR & TIR) menggunakan metode pembelajaran mesin = Classification and clasterization of rock minerals in field x based on spetral data (SWIR & TIR) using machine*. Universitas Indonesia.

Prandika Siregar, A., Priyadi Purba, D., Putri Pasaribu, J., Reza Bakara, K., & Willem Iskandar Pasar Medan Estate, J. V. (2023). Implementasi Algoritma Random Forest Dalam Klasifikasi Diagnosis Penyakit Stroke. *Jurnal Penelitian Rumpun Ilmu Teknik (JUPRIT)*, *2*(4), 155–164. https://doi.org/10.55606/juprit.v2i4.3039

Ran, X., Xue, L., Zhang, Y., Liu, Z., Sang, X., & He, J. (2019). Rock classification from field image patches analyzed using a deep convolutional neural network. *Mathematics*, *7*(8), 755. https://doi.org/10.3390/math7080755

Turlapati, V. P. K., & Prusty, M. R. (2020). Outlier-SMOTE: A refined oversampling technique for improved detection of COVID-19. *Intelligence-Based Medicine*, *3–4*. https://doi.org/10.1016/j.ibmed.2020.100023

Williams, H., Turner, F. J., & Gilbert, C. M. (1982). Petrography: an introduction to the study of rocks in thin sections. In *Petrography: an introduction to the study of rocks in thin sections*. W.H. Freeman Company. https://doi.org/10.5408/0022-1368-3.1.34