JURNAL ILMIAH EKONOMI DAN BISNIS, Vol.13, No.2, Desember 2020, pp. 67 - 75
p-ISSN : 1979-0155 (print)
e-ISSN : 2614-8870 (online)
http://journal.stekom.ac.id/index.php/E-Bisnis

Penerapan Algoritma Random Forest Untuk Menentukan Kualitas Anggur Merah

Riki Supriyadi[1], Windu Gata[2], Nurlaelatul Maulidah[3], Ahmad Fauzi[4]
[1,2,3]Ilmu Komputer, STMIK Nusa Mandiri
Jalan Margonda Raya No. 545, Pondok Cina, Depok, (021) 31908575
e-mail: 14002371@nusamandiri.ac.id[1], windu@nusamandiri.ac.id[2], 14002377@nusamandiri.ac.id[3]
[4]Sistem Informasi Akuntansi, Universitas Bina Sarana Infomatika
Jl. Letjen Sutoyo No.43, Kec. Banjarsari, Kota Surakarta, (0271) 854440, e-mail: ahmad.fzx@bsi.ac.id[4]

ARTICLE INFO

Article history:

Received 30 September 2020
Received in revised form 2 Oktober 2020
Accepted 10 Oktober 2020
Available online 22 Oktober 2020

ABSTRACT

In this study that was used as the object of research in classifying red wine based on the quality influenced by each red wine or red wine based on the content of each type of wine, from each attribute containing the composition in the wine seen which attributes most affect the quality of red wine, so that it will be known ingridents that can improve the quality of the wine, in this study was carried out by the application of Machine learning by comparing three algorithms of mining data that is , Decission Tree, Random Forest and Support Vector Machine (SVM), from the results of research that has been done by comparing the three algorithms, Random Forest produced the best accuracy among other algorithms that have been tested. Random Forest with accuracy results of 0.7468 makes this algorithm best used to classify the quality of red wine. And in the second order Decission Tree with accuracy results of 0.7031, while Support Vector Machine (SVM) get an accuracy result of 0.65. So in the research that has been done to classify the quality of red wine based on its composition Random Forest becomes the best algorithm to use..

Keywords : Red wine, Random Forest, Python

Abstrak

Pada penelitian ini yang dijadikan objek penelitian dalam mengklasifikasikan anggur merah berdasarkan kualitas yang dipengaruhi oleh setiap anggur merah atau *red wine* berdasarkan pada kandungan setiap jenis *wine*, dari setiap atribut yang berisi komposisi dalam *wine* dilihat atribut mana yang paling mempengaruhi kualitas dari *red wine*, sehingga akan diketahui *ingridents* yang bisa meningkatkan kualitas dari *wine* tersebut, dalam penelitian ini dilakukan dengan penerapan *Machine learning* dengan membandingkan tiga algoritma data mining yaitu, *Decission Tree*, *Random Forest* dan *Support Vector Machine* (SVM), dari hasil penelitian yang telah dilakukan dengan mebandingkan ketiga algoritma, dihasilkan *Random Forest* menghasilkan akurasi paling baik diantara algoritma lainya yang telah diuji. *Random Forest* dengan hasil akurasi 0.7468 menjadikan algoritma ini paling baik digunakan untuk

Received September 30, 2020; Revised Oktober 2, 2020; Accepted Oktober 22, 2020

p-ISSN : 2614-8870 e-ISSN : 2614-8870

memgklasifikasikan kualitas *red wine*. Dan diurutan kedua *Decission Tree* dengan hasil akurasi sebesar 0.7031, sedangkan *Support Vector Machine* (SVM) mendapatkan hasil akurasi sebesar 0.65. Jadi pada penelitian yang telah dilakukan untuk menngklasifikasikan kualitas *red wine* berdasarkan komposisinya *Random Forest* menjadi algoritma paling baik untuk digunakan.

Kata Kunci : *Red wine, Random Forest,* Python

1. PENDAHULUAN

Anggur adalah salah satu buah yang banyak dikonsumsi dan cukup populer di semua wilayah. Buah anggur biasanya dikonsumsi secara langsung atau juga dapat dibuat suatu produk seperti makanan dan minuman yang difermentasi dari buah anggur. *Wine* merupakan minuman beralkohol yang dibuat dari sari buah anggur. *Wine* dibuat dengan melalui fermentasi gula yang ada di dalam buah anggur. Jangka waktu yang dibutuhkan saat fermentasi bervariasi waktunya ada yang singkat dan ada yang membutuhkan waktu yang lama. Ada beberapa jenis minuman anggur yaitu, *Rose Wine, Sweet Wine, Sparkling Wine, Red wine, White Wine, dan Fortified Wine.*

Red wine atau anggur merah adalah minuman anggur yang berasal dari buah anggur yang berwarna merah atau hitam. Warna merah dihasilkan merupakan hasil diperoleh dari pencelupan kulit dan biji ke dalam sari buah yang telah diperas untuk difermentasi dalam jangka waktu yang bervariasi.

Di luar negeri *wine* menjadi minuman yang banyak dikonsumsi oleh berbagai lapisan masyarakat, khususnya negara-negara yang memiliki iklim dingin dan bersalju yang berguna untuk menghangatkan badan saat turun salju. *Wine* juga sering disediakan di acara-acara tertentu dalam merayakan sebuah acara. Anggur memiliki berbagai karakteristik seperti kepadatan, nilai pH, alkohol dan asam lainnya.

Dalam perkembanganya *wine* semakin bermacam varianya. Hal itu pula yang membuat *wine* di bagi berdasarkan kualitanya untuk memnetukan harga jual di pasaran. Kulaitas pada *wine* dipengaruhi oleh bebrapa faktor, contohnya komposisi yang terdapat di dalamnya. Untuk menentukan kualitas *wine* tentu harus ada ahli yang bertugas untuk mencicipi sampel dari minuman anggur tersebut.

Menentukan kualitas dari minuman anggur yang difermentasikan terkadang sulit jika bukan ahlinya, dan dalam hal ini tidak banyak orang yang memiliki kemampuan dalam menentukan kualitas dari minuman anggur fermentasi, karena dibutuhkan indera perasa yang lebih peka terhadap komposisi yang terkandung didalam produk salah satunya *wine*. Penilaian kualitas anggur merupakan salah satu elemen kunci dalam konteks ini dan penilaian ini dapat digunakan untuk sertifikasi. Jenis sertifikasi kualitas seperti itu membantu memastikan kualitas anggur di pasar [1]

Era digital saat ini, kita dapat mengukur tingkat kualitas *wine* yang ditawarkan dengan memanfaatkan berbagai macam *tools* yang ada untuk menghitung data mining seperti Matlab, Rapid Miner, Weka dan python untuk menghitung tingkat kualitas *wine.*

Machine Learning adalah salah satu metode yang sering diterapkan oleh banyak peneliti. *machine learning* diperkenalkan untuk membantu dalam meningkatkan kemampuan pendeteksian otomatis [2]. Metode *machine learning* bekerja secara otomatis yang efisien dan efektif model klasifikasi karena mereka mengadopsi campuran metode matematika dan pencarian dari ilmu komputer [3].

Dalam bidang *food and drink* data mining digunakan untuk mengklasifikasikan makanan dan minuman berdasarkan kualitas, jenis dan lainya untuk menentuan *range* harga jual yang akan ditetapkan di pasaran nantinya.seperti minuman aggur yang difermentsikan yang sering disebut dengan *wine,* dalam hal ini adalah *red wine* yang dihasilkan dari anggur merah. Data mining digunakan untuk mengklasifikasi kualitas dari dataset *red wine* yang kemudian hasilnya digunakan untuk menaganalisa atribut mana yang paling mempengaruhi dari kualitas *red wine*.

p-ISSN : 1979-0155 e-ISSN : 2614-8870■

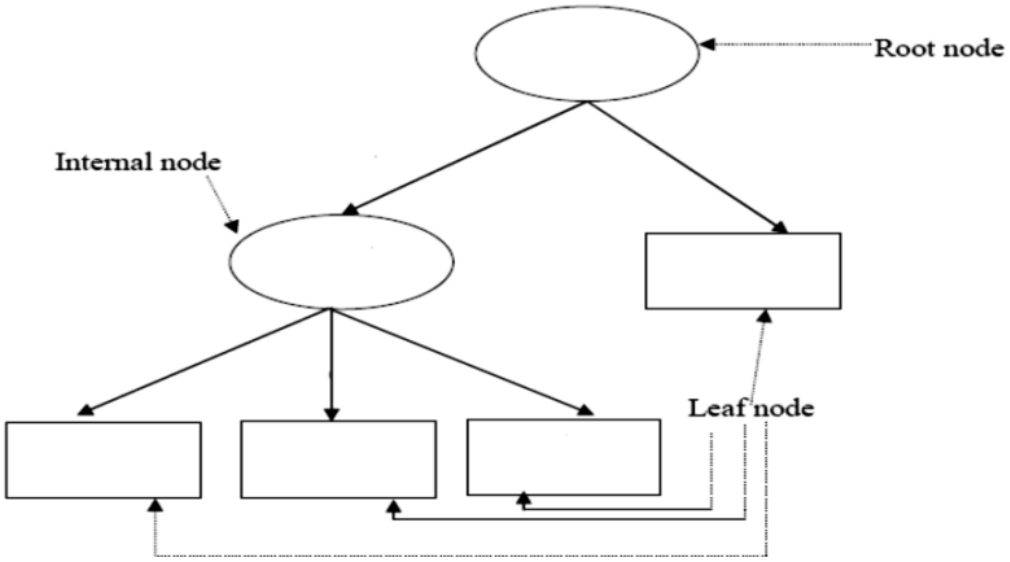
Dalam penelitian yang akan dilakukan terhadap dataset *Red wine Quality* dengan membandingkan tig aalgoritma yaitu, *Decission Tree*, *Random Forest* dan *Support Vector Machine* (SVM), untuk melakukan pengujian akan dilakukan dengan *tools* Python

2. TINJAUAN PUSTAKA

Pada penelitian ini akan memprediksi hasil klasifikasi dari tiga algoritma yang kemudian akan dibandingkan masilnya sertiap algoritma untuk menentukan algoritma yang paling baik tehadap pengujian dataset.

2.1. *Decission Tree*

Decission Tree merupakan metode penelitian yang paling sering digunakan untuk masalah klasifikasian. decision tree adalah sebuah struktur yang bisa digunakan untuk membagi kumpulan data besar menjadi himpunan-himpunan *record* yang lebih kecil melalui serangkaian aturan keputusan. Setiap simpul pada daun menandai label kelas. Simpul yang bukan simpul akhir terdiri dari simpul internal dan akar yang terdiri dari kondisi tes atribut pada sebagian *record* yang memiliki karakteristik yang berbeda. Simpul simpul internal dan akar ditandai dengan bentuk oval dan simpul daun ditandai dengan segi empat. [4]



Sumber : Muzakir, Wulandari (2016)

Gambar 1. Struktur *Decission Tree* [1]

2.2. *Random Forest*

Random Forest adalah pengembangan dari metode *Decision Tree* yang menggunakan beberapa *Decision Tree,* dimana setiap *Decision Tree* telah dilakukan pelatihan menggunakan sampel individu dan setiap atribut dipecah pada pohon yang dipilih antara atribut subset yang bersifat acak**.** *Random Forest* memiliki beberapa kelebihan, yaitu dapat meningkatkan hasil akurasi jika terdapat data yang hilang, dan untuk *resisting outliers*, serta efisien untuk penyimpanan sebuah data. Selain itu, *Random Forest* mempunyai proses seleksi fitur dimana mampu mengambil fitur terbaik sehingga dapat meningkatkan performa terhadap model klasifikasi. Dengan adanya seleksi fitur tentu *Random Forest* dapat bekerja pada *big data* dengan parameter yang kompleks secara efektif. [5]

2.3. *Support Vector Machine* (SVM)

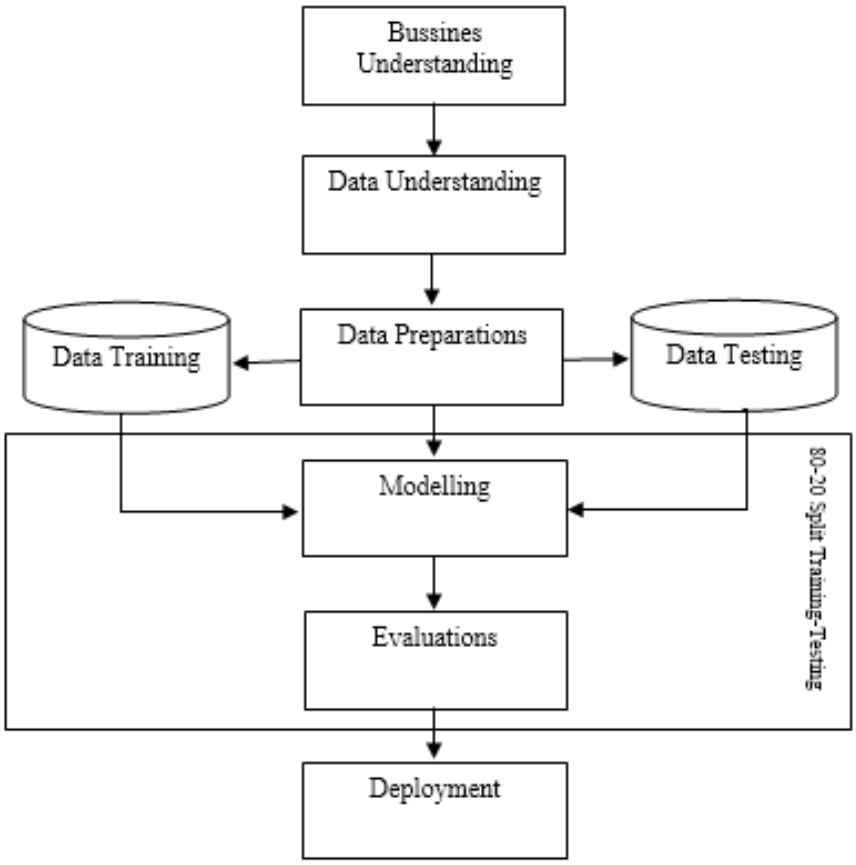
Algoritma *Support Vector Machine* (SVM) merupakan algoritma yang bisa melakukan prediksi berupa klasifikasi. *Support Vector Machine* (SVM) adalah salah satu metode klasifikasi yang mempunyai prinsip mencari *hyperplane* yang memiliki *margin* terbesar. *Hyperplane* merupakan suatu garis yang memisahkan sebuah data antar *class* atau kategori. Sedangkan margin yaitu jarak antara *hyperplane*

p-ISSN : 2614-8870 e-ISSN : 2614-8870

dengan data terdekat yang berada di masing-masing kategori. Data paling dekat dengan *hyperplane* disebut dengan *support vector* . Beberapa metode diusulkan agar SVM dapat dipergunakan untuk klasifikasi *multi-class* dengan kombinasi beberapa *binary classifier*. [6] *multiclass Support Vector Machine* (SVM) ditransformasikan ke beberapa *classifier binary*. Setiap *classifier binary* SVM dilatih menggunakan matriks data *training*, di mana setiap baris sesuai dengan fitur yang diekstraksi sebagai pengamatan dari *class*. [7]

3. METODOLOGI PENELITIAN

Pada penelitian ini dilakukan dengan beberapa tahapan, setiap tahapanya dapat dilihat pada gamabar 2.



Sumber : Penelitian (2020)

Gambar 2. Tahapan Penelitian [2]

3.1. *Business Understanding*

Tahap pertama yaitu *Business Understanding* pada tahap ini memahami kebutuhan atau tujuan dari dataset yang akan diteliti. Berikut tahapan dalam bussines understanding :

a. Menentukan tujuan bisnis, yaitu memprediksi kualitas *red wine* berdasarkan komposisi atau *ingridients* yang terkandung di dalam *red wine*.
b. Menilai situasi, beberapa produk *wine* memiliki kualitas yang berbeda dari kualitas buruk, sedang, dan bagus yang pengaruhi oleh kandungan bahan di dalamnya, jadi perlu mengetahui kandungan apa yang paling berpengaruh terhadap kualitas dari *red wine*.
c. Menentukan tujuan data mining, tujuan dari penelitian ini adalah meningkatkan pengetahuan tentang kualitas *wine* untuk memenuhi kebutuhan dan keinginan konsumen dan memberi petunjuk tentang kemungkinan, dan kesediaan konsumen untuk membeli anggur dengan campuran bahan-bahan tertentu serta memberikan keunggulan bagi produsen dibandingkan pesaing lainnya.

p-ISSN : 1979-0155 e-ISSN : 2614-8870■

3.2. *Data Understanding*

Dataset yang digunakan yaitu berasal dari kaggle tentang *Red wine* Quality dari anggur "Vinho Verde" Portugal. Pada dataset memilki *input* dan *output* yang menetapkan nilai kualitas mulai dari 3 sampai 8, dimana tiga berarti buruk, dan delapan berarti sangat baik. Ada 1.599 data sampel observasi dengan12 atribut yaitu fixed acidity, citric acid, volatileacidity, residual sugar, chlorides, free sulfur dioxide, total sulfur dioxide, density, pH, sulphates, alcohol dan qualit. Dalam penerapan penelitian ini menggunakan python.

Berikut rincian dari dataset *Red wine* quality :

```
wine.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1599 entries, 0 to 1598
Data columns (total 12 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   fixed acidity         1599 non-null   float64
 1   volatile acidity      1599 non-null   float64
 2   citric acid           1599 non-null   float64
 3   residual sugar        1599 non-null   float64
 4   chlorides             1599 non-null   float64
 5   free sulfur dioxide   1599 non-null   float64
 6   total sulfur dioxide  1599 non-null   float64
 7   density               1599 non-null   float64
 8   pH                    1599 non-null   float64
 9   sulphates             1599 non-null   float64
 10  alcohol               1599 non-null   float64
 11  quality               1599 non-null   int64
dtypes: float64(11), int64(1)
memory usage: 150.0 KB
```

Sumber : Hasil Penelitian (2020)

Gambar 3. Atribut Dataset [3]

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol	quality
0	7.4	0.70	0.00	1.9	0.076	11.0	34.0	0.9978	3.51	0.56	9.4	5
1	7.8	0.88	0.00	2.6	0.098	25.0	67.0	0.9968	3.20	0.68	9.8	5
2	7.8	0.76	0.04	2.3	0.092	15.0	54.0	0.9970	3.26	0.65	9.8	5
3	11.2	0.28	0.56	1.9	0.075	17.0	60.0	0.9980	3.16	0.58	9.8	6
4	7.4	0.70	0.00	1.9	0.076	11.0	34.0	0.9978	3.51	0.56	9.4	5

Sumber : Hasil Penelitian (2020)

Gambar 4. Sampel Dataset [4]

3.3. *Data Preparation*

Pada tahap ini data disiapkan untuk dilakukan proses pelatihan. Datanya sendiri terdiri dari 1599. Kemudian tahap pengolahan data yang bertujuan untuk membangun dataset akhir yang akan diproses pada tahap pemodelan. Tahapan pengolahan data mencakup pemilihan kelas, atribut-atribut data serta transformasi data kemudian dilakukan proses pembersihan dan regulasi data, pada proses ini mencoba untuk menghilangkan missing value dan lain sebagainya. Dari 12 atribut yang terdapat di dalam dataset, atribut yang dijadikan kelas yaitu Quality.

p-ISSN : 2614-8870 e-ISSN : 2614-8870

3.4. *Modelling*

Algoritma yang digunakan dalam data latih ini ada tiga yaitu *Decission Tree*, *Random Forest* dan *Support Vector Machine* (SVM) . Data yang akan digunakan untuk peramalan dibagi menjadi dua, yaitu data 80 training dan 20 data testing. Dengan algoritma menggunakan tiga algoritma bertujuan untuk mendapatkan model terbaik dengan membandingkan masing-masing hasil yang diperoleh untuk dapat melakukan prediksi dengan tingkat akurasi tertinggi terhadap kualitas dari *red wine* dan memprediksi kandungan apa yang paling berpengaruh besar terhadap kualitas *wine*, sehingga memberikan hasil yang optimal.

3.5. *Evaluation*

Pada tahap ini dilakukan evaluasi berdasarkan ketiga algoritma yaitu *Decission Tree*, *Random Forest* dan *Support Vector Machine* (SVM), kemudian dibuat perbandingan untuk memilih yang paling tepat berdasarkan nilai dari hasil accuracy, confusion matrix, AUC dan F1 score

3.6. *Deployment*

Setelah tahap evaluasi dilakukan dengan menilai secara detail dari hasil permodelan dan dilakukan pengimplentasian dari keseluruhan Model yang telah di rancang sehingga mendapatkan hasil sesuai yang diharapkan. Prediksi kualitas *red wine* berdasarkan kandungan yang ada di dalamnya dapat dikembangkan agar selanjutnya dapat menentukan harga jual yang sesuai dengan kualitas dari *wine*.

4. HASIL DAN PEMBAHASAN

Berikut merupakan hasil penelitian yang telah dilakukan dengan membandingkan tiga algoritma untuk memprediksi kualitas dari *red wine* :

4.1 Hasil Penelitian

Pada penelitian ini dilakukan dengan tiga algoritma yang berbeda untuk melakukan klasifikasi terhadap kualitas anggur merah, berikut hasil dari masing-masing algoritma :

4.1.1 *Decission Tree*

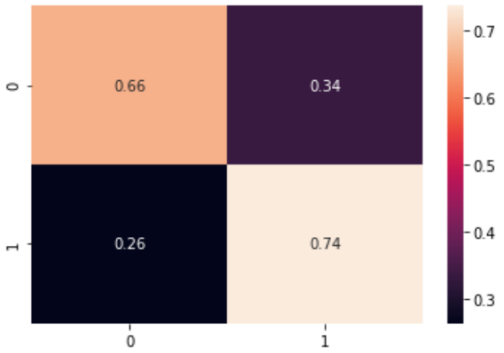
Pada pengujian yang dilakukan dengan algoritma *Decission Tree* mendapatkan hasil seperti pada tabel 1.

Tabel 1. Hasil *Decission Tree* [1]

Accuracy	0.7031
AUC	0.7000
F1 Score	0.7293

Sumber : Penelitian (2020)

Berdasarkan hasil penelitian yang telah dilakukan dengan algoritma *Decission Tree* menunjukan hasil akurasi sebesar 70,31 persen, dengan nilai Area Under Curve (AUC) 70,00 persen dan F1 Score 72,93 persen.



Sumber : Penelitian (2020)

Gambar 5. *Confussion matrix Decission Tree* [5]

Berikut nilai *confussion matrix* dari *Decission Tree*:
True Positif : bernilai 0.66
True Negatif : bernilai 0.74
False Positif : bernilai 0.34
False Negatif : bernilai 0.26

4.1.2 *Random Forest*

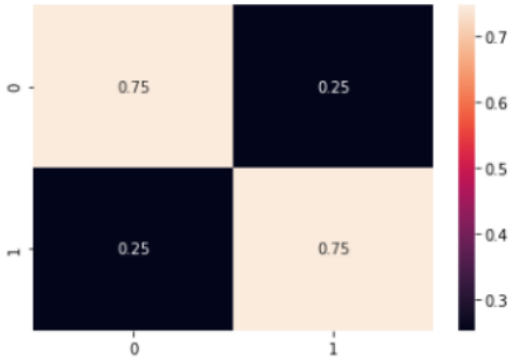
Pada pengujian yang dilakukan dengan algoritma *Random Forest* mendapatkan hasil seperti pada tabel 2.

Tabel 2. Hasil *Random Forest* [2]

Accuracy	0.7468
AUC	0.7468
F1 Score	0.7492

Sumber : Penelitian (2020)

Berdasarkan hasil penelitian yang telah dilakukan dengan algoritma *Random Forest* menunjukan hasil akurasi sebesar 74,68 persen, dengan nilai Area Under Curve (AUC) 74,68 persen dan F1 Score 74,92 persen.



Sumber : Penelitian (2020)

Gambar 6. *Confussion matrix Random Forest* [6]

Berikut nilai *confussion matrix* dari *Random Forest*:
True Positif : bernilai 0.75
True Negatif : bernilai 0.75
False Positif : bernilai 025
False Negatif : bernilai 0.25

4.1.3 *Support Vector Machine* (SVM)

Pada pengujian yang dilakukan dengan algoritma *Support Vector Machine* (SVM) mendapatkan hasil seperti pada tabel 3.

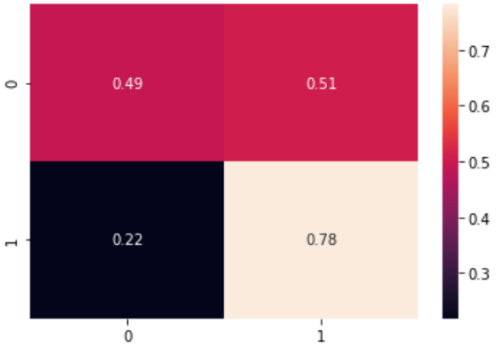
Tabel 3. Hasil SVM [3]

Accuracy	0.65
AUC	0.6373
F1 Score	0.7083

Sumber : Penelitian (2020)

Berdasarkan hasil penelitian yang telah dilakukan dengan algoritma *Support Vector Machine* (SVM) menunjulan hasil akurasi 65 persen, dengan nilai Area Under Curve (AUC) 63,73 persen dan F1 Score 70,83 persen.

p-ISSN : 2614-8870 e-ISSN : 2614-8870



Sumber : Penelitian (2020)

Gambar 7. *Confussion matrix Support Vector Machine* (SVM) [7]

Berikut nilai *confussion matrix* dari *Support Vector Machine* (SVM):
True Positif : bernilai 0.66
True Negatif : bernilai 0.74
False Positif : bernilai 0.34
False Negatif : bernilai 0.26

4.2 Hasil Perbandingan

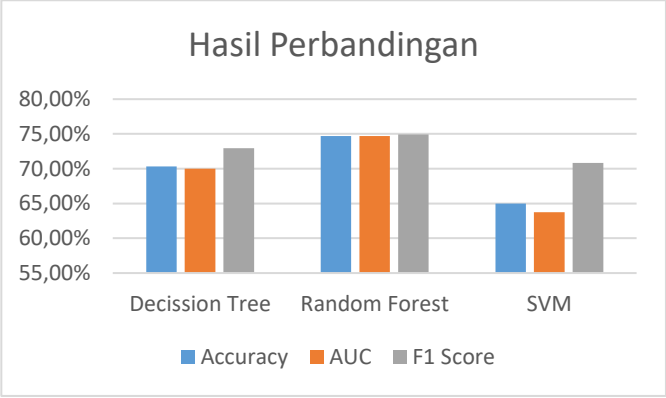
Setelah dilakukan pengujian satu per satu dari tiap algoritma kemudian hasil yang didapatkan dibandingkan untuk menentukan algoritma yang cocok untuk mengklasifikasikan kualitas *red wine*.
Berdasarkan hasil penelitian yang telah dilakukan berikut hasil perbandingan dari ketiga metode algoritma :

Tabel 4. Hasil Perbandingan [4]

Algoritma	Accuracy	AUC	F1 Score
Decission Tree	0.7031	0.7000	0.7293
Random Forest	0.7468	0.7468	0.7492
SVM	0.65	0.6373	0.7083

Sumber : Penelitian (2020)

Berdasarkan tabel 4. Dapat dilihat dari ketiga algoritma *machine learning* yaitu *Decission Tree*, *Random Forest* dan *Support Vector Machine* (SVM). Yang menunjukan hasil terbaik adalah pada algoritma *Random Forest* dibandingkan dengan algoritma lainya.



Sumber : Penelitian (2020)

Gambar 8. Hasil Perbandingan [8]

5. KESIMPULAN DAN SARAN

Penelitian yang telah dilakukan dengan membandingkan tiga algoritma data mining yaitu *Decission Tree*, *Random Forest* dan *Support Vector Machine* (SVM) untuk mengklasifikasikan kualitas dari red wine berdasarkan komposisi atau ingrideints yang terdapat didalamnya. Kemudian dibandingkan hasil dari setiap pengujian yang telah dilakukan didapatkan hasil *Decission Tree* dengan nialai akurasi 0.7031, niali AUC 0.7000 dan F1 Score 0.7293. Untuk *Random Forest* mendapatkan hasil akurasi 0.7468, AUC 0.7468 dan F1 Score sebesar 0.7492 dan untuk *Support Vector Machine* (SVM) mendapatkan hasil akurasi sebesar 0.65, AUC 0.6373 dan F1 Score sebesar 0.7083. berdasarkan hasil penelitian yang telah dilakukan maka dapat disimpulkan bahwa *Random Forest* merupakan algoritma yang bisa digunakan unuk melakukan klasifikasi pada dataset red wine quality berdasarkan ingridents atau komposisi yang terkandung di dalamnya. Untuk penelitian selanjutnya bisa dikembangkan lagi dengan metode yang berbeda untuk mendapatkan hasil yang lebih baik dan juga dapat memprediksi harga berdasarkan kualitas dari anggur merah.

DAFTAR PUSTAKA

[1] Y. Gupta, "Selection of important features and predicting wine quality using machine learning techniques," *Procedia Comput. Sci.*, vol. 125, pp. 305–312, 2018.

[2] A. Fauzi, R. Supriyadi, and N. Maulidah, "Deteksi Penyakit Kanker Payudara dengan Seleksi Fitur berbasis Principal Component Analysis dan Random Forest," vol. 2, no. 1, 2020.

[3] F. Thabtah, "Autism Spectrum Disorder Screening : Machine Learning Adaptation and DSM-5 Fulfillment," pp. 1–6, 2017.

[4] A. Muzakir and R. A. Wulandari, "Model Data Mining sebagai Prediksi Penyakit Hipertensi Kehamilan dengan Teknik Decision Tree," *Sci. J. Informatics*, vol. 3, no. 1, pp. 19–26, 2016.

[5] S. Devella, Y. Yohannes, and F. N. Rahmawati, "Implementasi Random Forest Untuk Klasifikasi Motif Songket Palembang Berdasarkan SIFT," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 7, no. 2, pp. 310–320, 2020.

[6] O. Somantri, S. Wiyono, and D. Dairoh, "Metode K-Means untuk Optimasi Klasifikasi Tema Tugas Akhir Mahasiswa Menggunakan Support Vector Machine (SVM)," *Sci. J. Informatics*, vol. 3, no. 1, pp. 34–45, 2016.

[7] H. N. Irmanda and Ria Astriratma, "Klasifikasi Jenis Pantun Dengan Metode Support Vector Machines (SVM)," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 4, no. 5, pp. 915–922, 2020.