

Article history

Received May 10, 2023

Accepted May 30, 2023

DETEKSI DINI SISWA BERMASALAH MENGGUNAKAN METODE RANDOM FOREST DI SMK MUHAMMADIYAH 3 BANJARMASIN

Frista Rizky Rinandi¹⁾, Hery Heryanto²⁾, Koes Wiyatmoko³⁾

¹ Informatika/STIMIK LIKMI Bandung

² Informatika / Institut Teknologi Harapan Bangsa Bandung

³ Teknik Informatika/Politeknik Negeri Banjarmasin

email: fristabjmkalsel@gmail.com, hery.hrynto@gmail.com, koesmoko@gmail.com

Abstract

Education is the basic capital for the state in developing productive and quality human resources. To produce a good quality education is also influenced by the behavior of children during the school period. Some of the variables that influence children's behavior include the family, environment, play and economic level. From some of these variables, the Random Forest method can be used to predict the success rate of children by processing data based on predetermined variables. To produce a prediction with a high level of accuracy, a method is also needed, namely the Confusion Matrix, to process variables which are input, an online based application is created using Python whose output immediately produces predictions of "Bermasalah" or "Tidak Bermasalah".

Keywords: *Data mining, random forest, children's behavior, confusion matrix prediction.*

Abstrak

Pendidikan adalah modal dasar bagi negara dalam membangun sumberdaya manusia kearah yang produktif dan berkualitas. Untuk menghasilkan pendidikan dengan kualitas yang baik juga dipengaruhi oleh perilaku anak selama dalam masa sekolah. Beberapa *variable* yang mempengaruhi perilaku anak diantaranya adalah keluarga lingkungan bermain dan tingkat ekonomi. Dari beberapa variabel tersebut Metode *Random Forest* dapat digunakan untuk memprediksi tingkat keberhasilan anak dengan mengolah data berdasarkan variabel-variabel yang telah ditentukan. Untuk menghasilkan sebuah prediksi yang tingkat akurasi tinggi maka juga diperlukan metode yaitu *Confusion Matrik*, untuk memproses variabel-variabel yang merupakan *input* maka dibuat aplikasi berbasis *online* menggunakan Python yang menghasilkan *output* langsung menghasilkan prediksi “ Bermasalah “ atau “Tidak Bermasalah”.

Kata Kunci: *Data Mining, Random Forest, Perilaku Anak, Prediksi Confusion Matrik.*

1. PENDAHULUAN

Kenakalan remaja merupakan perilaku yang bersifat negatif yang bertentangan dengan norma-norma sosial atau hukum di satu wilayah. Kenakalan ini dilakukan oleh remaja yang sedang dalam masa pertumbuhan untuk mencari identitas atau jati diri. Bentuk perilaku kenakalan itu seperti perkelahian, persekusi, perusakan properti, narkoba, perilaku seksual serta bentuk pelanggaran hukum lainnya. Yang menjadi keprihatinan dan perlu mendapatkan perhatian karena perilaku tersebut juga terjadi di sekolah. Kondisi ini tentunya perlu perhatian khusus karena menyangkut masa depan dunia pendidikan kita, karena perilaku siswa akan menjadi tolak ukur berhasil tidaknya praktik pendidikan di Indonesia sudah berjalan dengan baik atau belum [1]. Faktor faktor yang mempengaruhi perilaku remaja tersebut antara lain adalah kondisi keluarga, lingkungan, teman bermain sebaya. Kuatnya pengaruh dari teman sebaya yang memiliki sifat kecenderungan pada kenakalan remaja [2].

Penelitian tentang prediksi telah dilakukan beberapa peneliti seperti [3] dengan topik “Lulus Tepat Waktu” dan metode yang digunakan adalah *Random Forest*. Dalam penelitian dengan metode yang sama [4] mengangkat topik tentang “Prediksi Curah Hujan”, sementara mahasiswa Universitas Muhammadiyah Surakarta [5] membahas tentang penerimaan konsumen terhadap produk mobil,

Dari beberapa jurnal penelitian akurasi hasil pengolahan data dengan Metode *Random Forest* diatas 70%.

Berdasarkan beberapa uraian dan jurnal sebelumnya maka topik penelitian ini membahas tentang deteksi perilaku anak di Sekolah Menengah kejuruan Muhammadiyah III menggunakan metode *Random Forest* yang bertujuan untuk membantu memutuskan siswa yang bermasalah atau tidak dimasa depannya.

2. METODE PENELITIAN

Random Forest

Ruang lingkup kegiatan penelitian ini berfokus pada prediksi masa depan anak dengan melihat perilaku selama dalam pendidikan menggunakan metode *Random Forest*. Sedangkan obyek penelitiannya pada siswa di SMK Muhammadiyah 3 Banjarmasin. Dengan menerapkan beberapa metode seperti bootstrap aggregating dan pemilihan fitur secara acak. Kemudian dilakukan penarikan contoh yang acak

dengan pemulihan data. Pohon pada metode ini dibangun hingga mencapai ukuran yang maksimum tanpa adanya pemangkasan. Pada setiap simpul, pemilihan data *training* dilakukan dengan memilih secara acak [6]

Rational Unified Process (RUP)

Tahap dimana sistem dibangun untuk deteksi siswa bermasalah. Adapun langkahnya adalah sebagai berikut. Adapun langkahnya adalah sebagai berikut :

- a. *Inception*, tahap menentukan ruang lingkup lingkup pengembangan sistem yang dibutuhkan, melalui wawancara dan observasi sesuai kebutuhan penelitian.
- b. *Elaboration*, tahap melakukan pendalaman dari hasil inception agar lebih memahami permasalahan terhadap masalah yang sedang diteliti. Dua tahapan dalam Elaboration yaitu:
 - 1) Analisis, diperlukan untuk analisis permasalahan, persyaratan dan keputusan.
 - 2) Perancangan, diperlukan untuk perancangan sistem, dimana digunakan diagram UML yang meliputi diagram *use case*, *activity*, *class* dan *sequence* serta rancangan antar muka dan menggunakan tombol navigasi.
- c. *Construction*, tahap membangun sistem berdasarkan implementasi metode klasifikasi *random forest*, kemudian dilakukan uji coba dengan data yang didapatkan dilapangan. Kemudian dilakukan testing untuk menjamin kualitas dan kesesuaian perangkat lunak yang dibuat telah seperti yang diharapkan.
- d. *Transition*, tahap ini di perlukan untuk dokumentasi pada sistem untuk mempermudah pengguna dalam melakukan proses deteksi siswa.

Phyton dan Google Colab

Dalam membangun sistem *phyton* dipilih sebagai bahasa pemrograman, karena *Python* memiliki keterbacaan pada kode yang dibuat sangat tinggi, dengan begitu kode tersebut mudah di pahami oleh bahasa pemrograman dikarenakan memiliki library yang sangat banyak dan juga luas [7]. Sedangkan untuk mengoperasikan *phyton* menggunakan *Google Collab* untuk membuat *Integrated Environment Development (IDE)* secara online. DiGoogle Colab juga tersedia pustaka (*library*) yang dibutuhkan. Diantaranya , *TensorFlow*, *NumPy*, *Pandas*, dan pustaka lainnya [8].

Confusion Matrik

Metode ini diterapkan untuk proses perhitungan akurasi, *recall*, dan *precision* untuk mengevaluasi kemampuan pada sistem yang dibuat dengan hasil yang dicapai menemukan peringkat yang paling relevan, dan sebagai presentase dokumen di diambil yang benar-benar relevan pada *query*. *Recall* sebagai mengevaluasi kemampuan pada sistem untuk menemukan hasil pada semua item yang relevan dari dokumen dan hasil tersebut didefinisikan untuk presentase dokumen seberapa relevan terhadap *query*. *Accuracy* ialah perbandingan dari kasus yang diidentifikasi nilai benar dengan jumlah seluruh kasus [9].

	yes	no	Total
Actual	Yes	TP	FN
Calss	No	FP	TN
	Total	P'	N'
			P + N

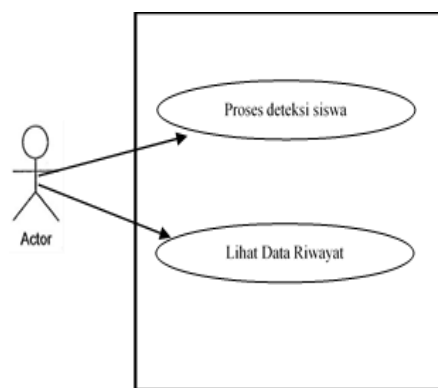
Gambar 1. Confusion Matrix

Keterangan Gambar Confusion Matrix:

- a. TP (*True Positive*) adalah data yang memiliki nilai keasliannya positif dan nilai dari hasil prediksi positif.
- b. FP (*False Positive*) adalah data yang memiliki nilai keasliannya negatif dan nilai dari hasil prediksi positif.
- c. FN (*True Negative*) adalah data yang memiliki nilai keasliannya positif namun dari hasil prediksi negative.
- d. TN (*True Negativ*) adalah data yang memiliki nilai keasliannya negative namun nilai hasil dari prediksi negative

Unified Modelling Language (UML)

Use Case Diagram pemodelan visual yang digunakan untuk perancangan sistem berorientasi objek.



Gambar 2. Use Case Sistem Deteksi random forest

Tabel 1. Deskripsi Use Case Pengguna

ID	Nama	Deskripsi
UC-01	Proses Deteksi Siswa	Pengguna dapat melakukan deteksi siswa bermasalah
UC-02	Lihat Data Riwayat	Pengguna dapat melihat riwayat dari hasil deteksi siswa bermasalah yang pernah dilakukan

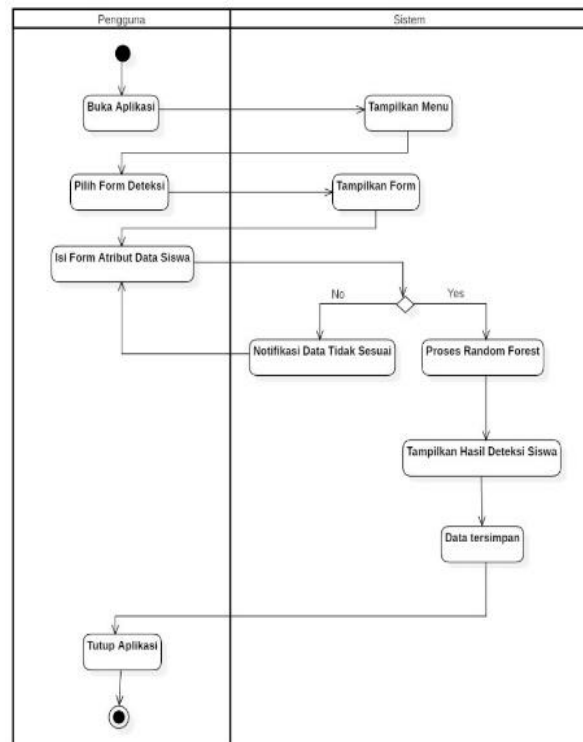
Tabel 2. Use Case Deteksi Siswa

Aksi Aktor	Reaksi Sistem
1. Buka aplikasi	
	2. Tampilkan menu
3. Pilih form deteksi	
	4. Tampilkan form
5. Isi form atribut data siswa	
	6. Proses perhitungan algoritma random forest
	7. Tampilkan hasil deteksi siswa
	8. Data tersimpan
9. tutup aplikasi	

Tabel 3. Use Case Lihat Data Riwayat

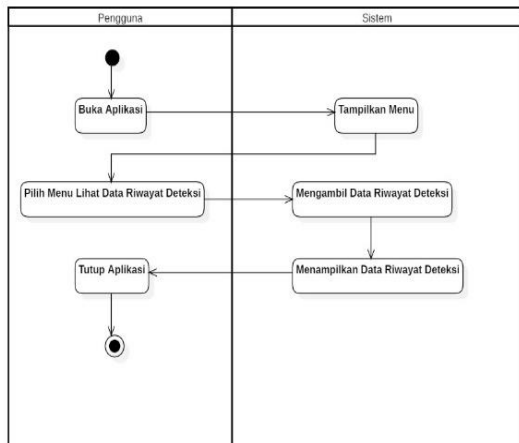
Aksi Aktor	Reaksi Sistem
1. Buka aplikasi	
	2. Tampilkan menu
3. pilih lihat data riwayat deteksi	
	4. Mengambil data riwayat deteksi
	5. menampilkan data riwayat deteksi
6. tutup aplikasi	

Activity Diagram digunakan untuk menggambarkan pemodelan perangkat lunak yang akan dibangun untuk sistem.



Gambar 3. Diagram Activity Deteksi Siswa

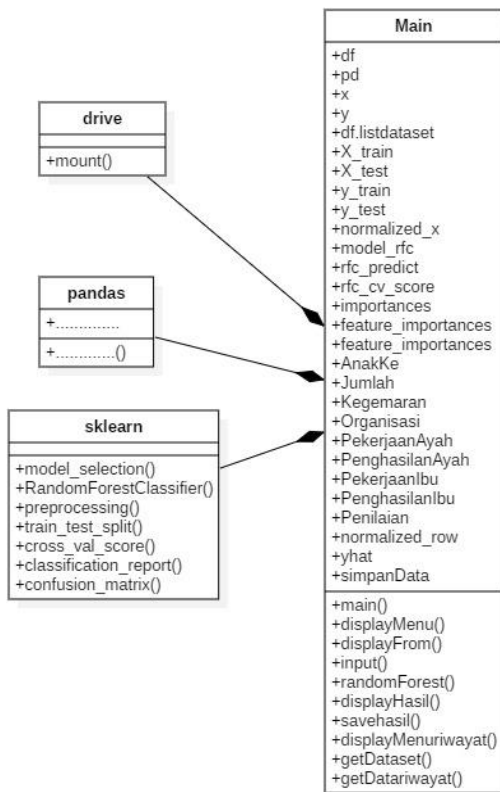
Pada Gambar 3 menunjukkan alur berjalannya aplikasi deteksi siswa bermasalah.



Gambar 4. Lihat Data Riwayat Deteksi

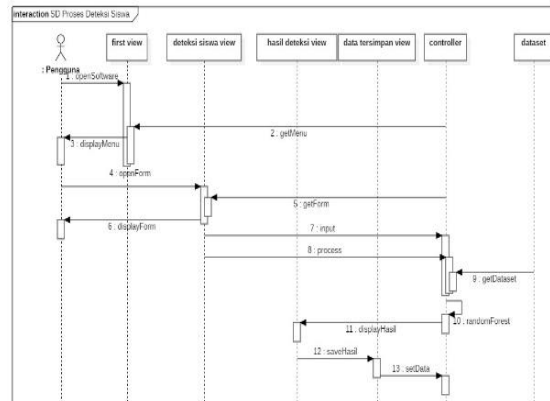
Pada Gambar 4 merupakan alur pengguna untuk proses lihat data riwayat.

Class diagram digunakan untuk rancangan dan variabel yang di perlukan pada sistem.

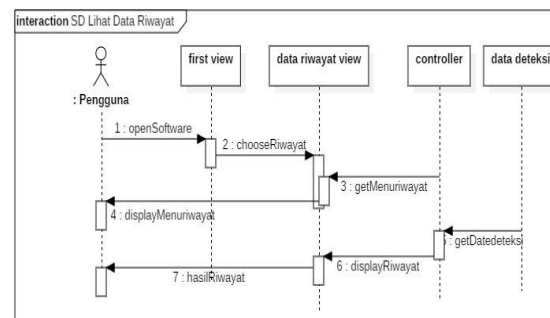


Gambar 5. Diagram Class Deteksi Siswa

Sequence diagram dari proses deteksi siswa dan lihat data riwayat. Dengan rancangan alur data yang disesuaikan oleh kebutuhan untuk sistem yang dibuat.

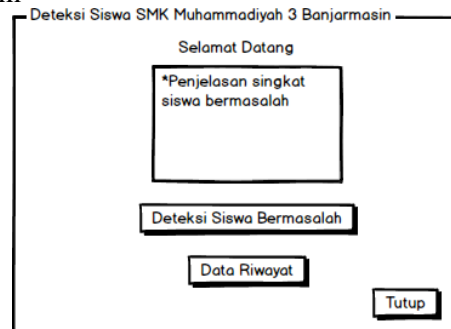


Gambar 6. Sequence Diagram Deteksi Siswa

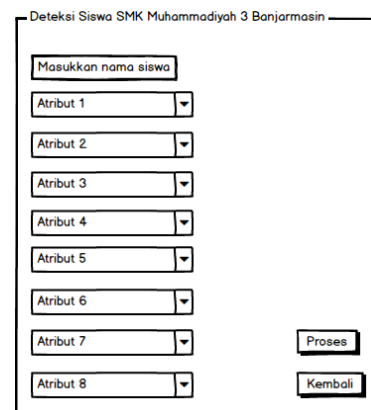


Gambar 7. Sequence Diagram Lihat Data Riwayat

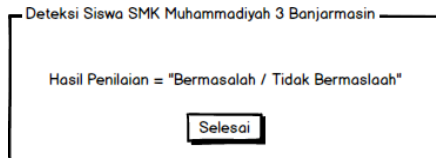
Interface merupakan rancangan antar muka untuk mempermudah pengguna dalam mengoperasikan sistem



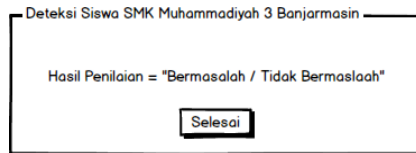
Gambar 8. Tampilan awal sistem



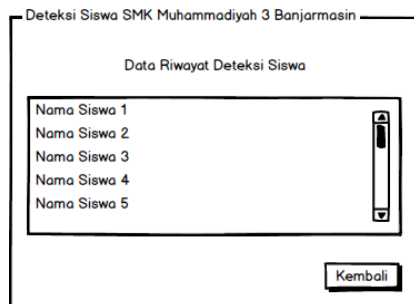
Gambar 9. Tampilan Deteksi Siswa



Gambar 10. Tampilan Dialog Hasil Deteksi



Gambar 11. Data Dialog Tersimpan



Gambar 12 Tampilan Data Riwayat Hasil Deteksi

3. HASIL DAN PEMBAHASAN

Proses uji coba sistem membutuhkan spesifikasi perangkat keras sebagai berikut :

- Processor Intel Core i7-8750H CPU @ 2.20GHz (12 CPus).*
- RAM 16GB.*
- Harddisk SSHD 1TB.*
- SSD Nvme M.2 120GB.*

Sedangkan spesifikasi kebutuhan perangkat lunak sebagai berikut :

- Windows 10 Home Single Language 64-bit.*
- Integrated Environment Development (IDE) Google Colab.*

Sedangkan parameter yang digunakan untuk mengukur adalah ;

- Organisasi, parameter ini digunakan untuk mengukur kemampuan dalam mengurus sesuai diluar kepentingan dirinya.

Tabel 3 Klasifikasi Parameter Organisasi.

Kode	Keterangan
0	Tidak Ada
1	Ada

- Kegemaran, parameter ini digunakan untuk mengukur emosi sesuai dengan hoby/kegemaran mereka karena dengan

sarana itu anak menyalurkan ekspresi dan melampiaskan emosi.

Tabel 4 Klasifikasi Parameter Kegemaran

Kode	Keterangan
0	Tidak ada
1	Olahraga
2	Pengetahuan dan Teknologi
3	Seni dan Kreasi

- Urutan Anak, parameter ini mempengaruhi perlakuan orang tua terhadap anak. Anak pertama cenderung dididik untuk menjadi contoh untuk saudaranya jika mempunyai adik. Anak bungsu cenderung bersifat manja dan banyak minta perhatian. Anak tengah seringkali dituntut dalam bersikap untuk dewasa dikarenakan untuk menghormati kakak ataupun adiknya.
- Jumlah Saudara, parameter berikut ini menjadi faktor-faktor yang mempengaruhi hubungan saudara kandung dan kedua orang tuanya. Hal ini berpengaruh terhadap kedisiplinan anak yang membentuk pola perilakunya..
- Penghasilan Orangtua , parameter ini mempengaruhi tingkat kenakalan oleh siswa. Dari data yang di kumpulkan, tingkat sosial ekonomi rendah sering berkelahi, membolos sekolah, mencuri, merokok, tawuran. Sedangkan siswa dari sosial ekonomi tinggi sering berjudi, menonton film porno, melakukan seks bebas dan mengkonsumsi obat terlarang.

Tabel 5 Klasifikasi Parameter Penghasilan Orang Tua

Kode	Keterangan
1	0-1.500.000
2	1.500.001-3.000.000
3	3.000.001-4.500.000
4	>4.500.000

- Pekerjaan Orangtua, parameter ini mengukur tingkat perhatian orang tua ketika dirumah. Orangtua lebih banyak waktu diluar rumah karena bekerja akan berbeda ketika orang tua sering dirumah. Anak cenderung mencari perhatian orang lain sehingga tidak terkontrol.

Tabel 6 Klasifikasi Parameter Pekerjaan Orang Tua

Kode	Keterangan
0	Tidak ada
1	Ada

- Pengujian Hasil
Pada tabel 7 dibawah ini merupakan input data yang akan di uji sistem.

Tabel 7 Pengujian Input Sistem

Parameter	Input
Anak ke-	3
Jumlah Saudara	3
Kegemaran	1 (Olahraga)
Organisasi	0 (Tidak Ada)
Pekerjaan Ayah	1 (Ada)
Penghasilan Ayah	4 (>Rp. 4.500.000)
Pekerjaan Ibu	1 (Ada)
Penghasilan Ibu	1 (Rp. 0 – 1.500.000)

Pengujian tersebut menghasilkan keputusan “Tidak Bermasalah”, dimana secara detail terlihat pada gambar dengan hasil laporan klasifikasi pada gambar 13

```

=== Confusion Matrix ===
[[424  19]
 [144  6]]

=== Classification Report ===
      precision    recall  f1-score   support

     0       0.75      0.96      0.84       443
     1       0.24      0.04      0.07       150

 accuracy          0.73       593
 macro avg          0.49      0.50      0.45       593
 weighted avg          0.62      0.73      0.64       593
    
```

Gambar 13. Hasil Pengujian inputan Tabel 7.

Penjelasan proses pengujian sistem :

- Data sebanyak 33% dari dataset data yang diacak sebanyak 66 kali.
- Nilai *confusion matrix* yang mengukur performa masalah klasifikasi, keluaran dapat memiliki dua kelas atau lebih. *Confusion Matrix* memiliki 4 tabel kombinasi berbeda dari suatu nilai prediksi dan nilai aktual. Ada empat istilah dari *confusion matrix* yang merupakan representasi dari hasil proses klasifikasi yaitu (1) *True positive* adalah data yang memiliki nilai keasliannya positif dan nilai dari hasil prediksi positif, (2) *false positive* adalah data yang memiliki nilai keasliannya negatif dan nilai dari hasil prediksi positif, (3) *true negative* adalah data yang memiliki nilai prediksi negatif dan (4) *false negative* adalah data yang memiliki nilai keasliannya negatif namun nilai hasil dari prediksi positif namun dari hasil prediksi negative.

Tabel 8 Confusion Matrix

Actual Class \ Predicted Class	P	N
	P	424
N	144	6

Precision merupakan prediksi benar positif yang dibandingkan dari data keseluruhan hasil prediksi positif. *Precision* menjawab pertanyaan “Berapa persen siswa yang tidak bermasalah dari keseluruhan siswa tidak

bermasalah?” .Untuk mendapatkan nilai *precision* didapatkan dari hasil perhitungan sebagai berikut :

$$\begin{aligned}
 \text{Precision} &= (\text{TP}) / (\text{TP}+\text{FP}) \\
 &= (424) / (424+144) \\
 &= 0,746479
 \end{aligned}$$

Recall merupakan prediksi benar positif yang dibandingkan dari data keseluruhan hasil data yang bernilai benar positif. *Recall* menjawab pertanyaan “ Berapa persen mahasiswa yang di prediksi bermasalah dibandingkan keseluruhan siswa yang sebenarnya tidak bermasalah. Untuk mendapatkan nilai *Recall* didapatkan dari hasil perhitungan sebagai berikut :

$$\begin{aligned}
 \text{Recall} &= (\text{TP}) / (\text{TP}+\text{FN}) \\
 &= (424) / (424+19) \\
 &= 0,957111
 \end{aligned}$$

F1 Score ialah perbandingan dari nilai rata-rata dari hasil nilai *precision* dan *recall* yang dibobotkan. Untuk mendapatkan nilai *F1 Score* didapatkan dari hasil perhitungan sebagai berikut :

$$\begin{aligned}
 \text{F1 Score} &= 2 * (\text{Recall} * \text{Precision}) / (\text{Recall} + \text{Precision}) \\
 &= 2 * (0,957589 * 0,748691) / (0,957589 + 0,748691) \\
 &= 0,838773
 \end{aligned}$$

Untuk mendapatkan nilai *Support* didapatkan dari hasil perhitungan sebagai berikut:

$$\begin{aligned}
 \text{Support} &= \text{TP} + \text{FN} \\
 &= 443
 \end{aligned}$$

Hasil dari perhitungan tersebut, maka didapatkan nilai *precision* 0,746479, *recall* 0,957111, *F1 Score* 0,838773 dan *support* 443.

4. PENUTUP

Kesimpulan

Dari hasil uji coba yang dilakukan dihasilkan perhitungan akurasi yang membuktikan bahwa variabel yang ditentukan serta metode yang digunakan mempunyai tingkat akurasi yang tinggi yaitu mendapatkan nilai dari *confusion matrix* *True positive* 424, *false positive* 144, *false negative* 19 dan *true negative* 6. Kemudian mendapatkan hasil nilai *precision* 0,746479, *recall* 0,957111, *F1 Score* 0,838773 dan *support* 443.

Saran

Penelitian ini kedepan sistem dapat dikembangkan dalam bentuk desktop sehingga tidak perlu koneksi ke internet. Selain itu sistem juga dapat dikembangkan berbasis mobile mengingat sebagian besar penduduk dunia saat

ini menggunakan smartphone untuk alat mencari informasi dan komunikasi.

5. REFERENSI

- [1] Nikmah Rahmawati, "KENAKALAN REMAJA DAN KEDISIPLINAN: Perspektif Psikologi dan Islam," vol. 11, no. April, pp. 267–288, 2016.
- [2] M. R. N., Nindya P., "Hubungan antara Kekerasan Emosional Pada Anak terhadap Kecenderungan Kenakalan Remaja," *Jurnal Psikologi Klinis dan Kesehatan Mental*, vol. 1, no. 02, pp. 1–9, 2012, [Online]. Available: [http://journal.unair.ac.id/filerPDF/110810221_ringkasan\(1\)_FIX_PKM.pdf](http://journal.unair.ac.id/filerPDF/110810221_ringkasan(1)_FIX_PKM.pdf)
- [3] I. M. Budi Adnyana, "Prediksi Lama Studi Mahasiswa Dengan Metode Random Forest (Studi Kasus: Stikom Bali)," *CSRID (Computer Science Research and Its Development Journal)*, vol. 8, no. 3, pp. 201–208, 2016, doi: 10.22303/csrid.8.3.2016.201-208.
- [4] A. Primajaya and B. N. Sari, "Random Forest Algorithm for Prediction of Precipitation," *Indonesian Journal of Artificial Intelligence and Data Mining*, vol. 1, no. 1, p. 27, 2018, doi: 10.24014/ijaidm.v1i1.4903.
- [5] Y. S. Nugroho, "Sistem Klasifikasi Variabel Tingkat Penerimaan Konsumen Terhadap Mobil Menggunakan Metode Random Forest," *Jurnal Teknik Elektro*, vol. 9, no. 1, pp. 24–29, 2017, doi: 10.15294/jte.v9i1.10452.
- [6] N. K. Dewi, U. D. Syafitri, S. Y. Mulyadi, M. D. Statistika, and D. Statistika, "Penerapan Metode Random Forest Dalam Driver Analysis," *Forum Statistika Dan Komputasi*, vol. 16, no. 1, pp. 35–43, 2011.
- [7] T. M. Kadarina and M. H. Ibnu Fajar, "Pengenalan Bahasa Pemrograman Python Menggunakan Aplikasi Games Untuk Siswa/I Di Wilayah Kembangan Utara," *Jurnal Abdi Masyarakat (JAM)*, vol. 5, no. 1, p. 11, 2019, doi: 10.22441/jam.2019.v5.i1.003.
- [8] R. T. Handayanto and H. Herlawati, "Prediksi Kelas Jamak dengan Deep Learning Berbasis Graphics Processing Units," *Jurnal Kajian Ilmiah*, vol. 20, no. 1, pp. 67–76, 2020, doi: 10.31599/jki.v20i1.71.
- [9] A.- Arini, L. K. Wardhani, and D.- Octaviano, "Perbandingan Seleksi Fitur Term Frequency & Tri-Gram Character Menggunakan Algoritma Naïve Bayes Classifier (Nbc) Pada Tweet Hashtag #2019gantipresiden," *Kilat*, vol. 9, no. 1, pp. 103–114, 2020, doi: 10.33322/kilat.v9i1.878.