INFOMATEK: Jurnal Informatika, Manajemen dan Teknologi
Volume 24 Nomor 1 Juni 2022

PREDIKSI OTORISASI PENGGUNA SISTEM BERKAS PADA ALGORITMA KLASIFIKASI DENGAN TEKNIK NAÏVE BAYES

Agus Pamuji*

Program Studi Bimbingan Konseling Islam, IAIN Syekh Nurjati Cirebon, Indonesia

Abstrak: Keamanan teknologi informasi saat ini menjadi tuntutan bagi setiap institusi. Salah satu bagian terpenting dalam keamanan IT adalah ketika melindungi sumber daya yang di-akses oleh beberapa pengguna. Beberapa kelemahan yang ditemukan adanya sistem yang mengizinkan beberapa pengguna untuk berbagi data sehingga berpotensi adanya potensi otorisasi pengguna yang berlebihan. Dalam penelitian ini akan dilakukan investi-gasi terhadap 162 log data aktifitas yang memuat data training dan data testing. Adapun dalam konsep analisisnya adalah menggunakan data mining dengan metode klasifikasi dan teknik Naïve Bayes. Hasil menunjukan metode klasifikasi dengan teknik Naïve Bayes cukup efektif untuk memprediksi adanya anomali otorisasi pengguna file. Selan-jutnya sudah dibuktikan melalui kinerjanya dengan menggunakan confusion matrix yang mencapai 89% dalam nilai akurasinya. Dengan demikian, data mining akan menjadi se-buah konsep dalam pengembangan dan analisis terhadap sistem keamanan terutama sis-tem berkas yang mengakomodasi manajemen pengguna untuk mengizinkan akses sum-ber daya.

Kata kunci: klasifikasi, Naïve Bayes, Otorisasi, pengguna, sistem berkas

I. PENDAHULUAN

Kondisi saat ini, sudah banyak diketahui oleh orang bahwa teknologi (Gunawan, 2019) telah hadir tanpa mempertimbangkan berbagai perubahan (Putri, 2021) . Perubahan yang sudah dan sedang terjadi saat ini dianggap signifikan terutama terhadap sistem keamanan berkas (Irmayani, 2021). Keamanan dalam komputer itu sangat penting walaupun sebagian pengguna masih menganggap memiliki tingkat resiko rendah (Chazar, 2020). Kenyataannya, apabila ada kerusakan, kehilangan data yang diakibatkan penyusup maka menjadikan sistem keamanan tidak terjamin. Dengan demikian, sistem berkas dalam berbagai atribut dan kondisinya dilengkapi dengan kontrol akses bagi setiap pengguna pada lingkungan internal dan eksternal (Das, 2018). Sebagai tambahannya, dengan adanya kontrol akses akan membuat pengguna merasa aman dan nyaman ketika berinteraksi dengan sistem (Dulhare, 2018).

Pada konteks ini, kontrol akses yang dimiliki oleh setiap pengguna juga memungkinkan dapat mengakses sumber daya yang sesuai dapat dianggap memiliki masalah. Permasalahan utama adalah pemuatan informasi yang sensitif pada sistem berkas. Berdasarkan observasi yang sudah dilakukan banyak ditemukan pengaksesan sistem berkas tidak sah (Supriyatna & Mustika, 2018). Penyebabnya adalah data dan juga file sudah terlalu banyak dibagikan kepada beberapa

*) jurnal.agus.pamuji@gmail.com

Diterima: 22 Oktober 2021
Direvisi: 3 Januari 2022
Disetujui: 6 Juni 2022
DOI: 10.23969/infomatek.v24i1.4604

pengguna tanpa adanya batasa tertentu (Fadlil et.al, 2017). Selain itu, metode untuk memproteksi dalam sistem berbagi melibatkan pengguna diperlukan dan kontrol akses yang mutlak (Degirmenci, 2019). Tahap selanjutnya, pengguna menggunakan hak izin yang bersifat relatif untuk mengakses sumber daya sesuai dengan identitas secara personal, disertai hak yang bersangkutan, dan wewenang pengguna (Parlina, 2019).

Berdasarkan literatur yang sudah ditelusuri, masalah izin pada penggguna yang bersifat berlebihan masih sedikit dalam hal publikasi selain penelitian didalamnya (Cai, 2019). Dengan contoh yang ada, hak memiliki akses sistem berkas berlebihan yang diterima oleh seorang pengguna dapat berpeluang meningkatkan risiko kehilangan data, anomali data, tindakan kriminal pencurian data, dan modifikasi bahkan sampai tindakan fabrikasi data (Rifqo & Wijaya, 2017).

Selanjutnya, permasalahan lebih mendalam menunjukan ketika perusahaan sering mengalami sejumlah kasus perizinan sistem berkas yang berlebihan terhadap data berpeluang untuk melakukan kegiatan penyalahgunaan. Adapun tujuan dari penelitian ini berespektasi dengan mencoba mengajukan pendekatan dengan menggunakan konsep data minin khususnya pada algoritma klasifikasi dengan teknik Naive Bayes (Taylor et.al, 2020). Secara konseptual data mining memiliki kemampuan untuk memprediksi ancaman perizinan yang berlebihan tercakup pada sistem keamanan berkas komputer (Laksono et.al, 2019). Segala sesuatu yang berkaitan dengan data akan memberikan banyak perhatian yang digunakan untuk menginvestigasi secara kritis kemungkinan penyalahgunaan hak dan izin disebabkan pada perizinan yang berlebihan (Wongkar & Angdresey, 2019).

Dengan demikian, data mining yang merupakan studi memiliki dua model dari berbagai literatur (Wahyuningish, 2018). Pertama, model prediktif didefinisikan memiliki model yang memiliki kinerja dengan prediksi hasil tertentu. Kedua, model deskriptif dalam konteks ini diartikan sebagai model yang dibuat dengan memberikan pemahaman yang lebih tentang data tanpa variabel tertentu dengan menggunakan teknik analisis seperti analisis faktor dan analisis klaster (Wood, et.al, 2019). Dengan latar belakang ini, konsep data mining memiliki banyak metode termasuk algoritma dengan kemampuan untuk data besar mengenai data besar (Aridas et.al, 2019). Beberapa metode dalam data mining dapat dikelompokkan menjadi estimasi, prediksi, klasifikasi, clustering, dan asosiasi. Pada konsep penelitian ini dapat menggunakan satu kelompok yaitu klasifikasi menggunakan metode Naive Bayes, K-Nearest Neighbor, Decission Tree, Random Forest, dan algoritma C4.5 (Peling et.al, 2017).

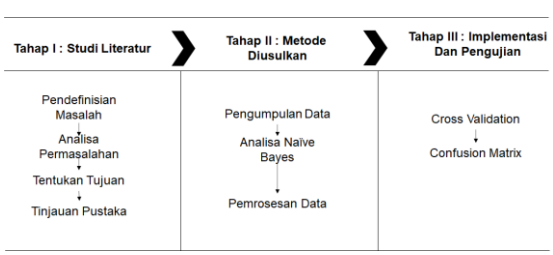
Motivasi utama yang terkait dengan penelitian ini telah dilakukan dengan tujuan untuk menghasilkan kerangka kerja yang digunakan untuk menganalisis adanya anomali data pada sistem berkas secara keseluruhan. Meskipun teknik yang ada saat menyentuh dengan izin berlebihan misalnya untuk metode tradisional, teknik kriptografi dan pengaturan akses pengguna individu. Sistem berkas yang sudah diketahui berisi informasi tentang akun pengguna dan memperoleh hak akses ke data dan riwayat aktivitas apa yang dilakukan (Rasjid & Setiawan, 2017). Untuk mencapai pada penelitian ini, dengan konseptualisme data mining disertai dengan metode dan

algoritma dalam klasifikasi data untuk memprediksi dan mengklasifikasikan sebagai identifikasi kelas pada hak akses dan izin (Gunduz & Das, 2019).

Kajian keamanan sistem berkas, khususnya terkait hak akses dan perizinan, kurang mendapat perhatian dan belum menjelaskan bagaimana cara memprediksi bahaya dan ancaman yang akan datang (Fadlan et.al, 2018). Oleh karena itu, beberapa peneliti hanya menyediakan metode dan algoritma yang dilengkapi dalam menangani masalah yang ada, tidak menganalisisnya sebelum terjadi. Berikut ini adalah beberapa kontribusi utama dari metode dan pendekatan penelitian yang diusulkan (Syukri et.al, 2017). Kontribusi yang dilakukan antara lain upaya untuk mengimprovisasi teknik dan metode keamanan sistem berkas sharing, memprediksi ancaman izin yang berlebihan, mengidentifikasi potensi izin yang berlebihan melalui klasifikasi, dan meningkatkan efisiensi dan efektivitas (Prackevicuus & Marchinkevičius, 2017).

II. METODOLOGI

Tahap metode penelitian, disajikan usulan kerangka kerja dalam memprediksi probabilitas adanya otorisasi pengaksesan sistem berkas. Ada tiga tahapan yang diusulkan yang dapat disajikan pada Gambar 1.



Gambar 1. Kerangka Kerja Penelitian

Tahap pertama yaitu studi literatur yang mencakup pendefinisian masalah, analisis permasalahan, menentukan tujuam dan tinjauan pustaka (Imandasari et.al, 2019). Kedua, metode yang diusulkan meliputi pengumpulan data, menganalisis dengan Naïve Bayes dan melakukan pemrosesan data. Ketiga, tahap implementasi dan pengujian meliputi melakukan validasi model dengan menggunakan cross validasi dan Confusion Matrix (Annur, 2018).

2.1. Studi Literatur

Pendefinisian permasalahan

Pada tahap ini akan dilakukan tinjauan terhadap sistem untuk mengobservasi lebih dalam dan telusuri masalah yang ada dalam sistem keamanan terutama pada sistem berkas. Tahap ini adalah langkah pertama untuk menentukan rumusan masalah penelitian. Adapun definisi pada konteks ini adalah otoritas pemberian izin akses pengguna banyak menimbulkan anomali yang berpotensi pada kehilangan, pencurian dan penyalahgunaan data pada sistem berkas. Definisi kedua, perlu dilakukan analisis probabilitas dalam memprediksi adanya potensi otorisasi pengaksesan sistem berkas yang tidak sah (Slamet et.al, 2018). Tahap pendefinisian ini terdapat empat variabel yang dapat diobservasi yaitu Tipe otorisasi, Peringkat Pengguna, Tipe Akun Pengguna, dan Tipe Pengguna.

Selanjutnya ada dua kelas klasifikasi yaitu “Ya” mendeskripsikan adanya potensi anomali data pada aktifitas otorisasi sistem berkas dan “Tidak” mendeskripsikan adanya potensi anomali data pada aktifitas otorisasi sistem berkas.

Analisis Permasalahan.

Masalah yang ditemukan kemudian akan dianalisis. Langkah dalam proses analisis masalah adalah langkah untuk memahami masalah yang telah ditentukan. Oleh sebab itu, menganalisis masalah yang telah ditentukan, yaitu diharapkan masalah dapat dipahami dengan baik. Dalam bagian ini sudah jelas bahwa perlu dilakukan prediksi dan juga probabilitas dalam mengurangi resiko adanya anomali otorisasi sistem berkas terutama pemberian hak akses pengguna.

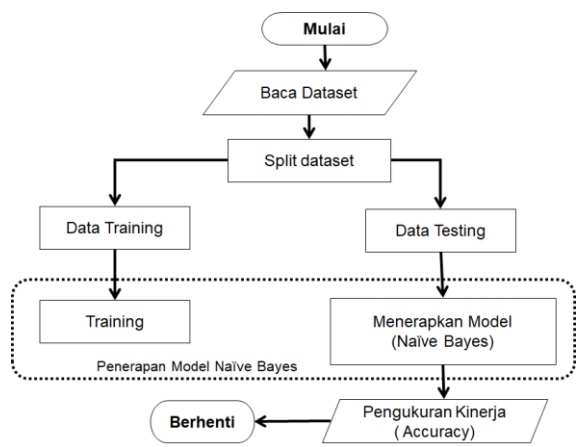
Menentukan Tujuan.
Berdasarkan pemahaman yang berasal dari pendefinisian masalah yang ada dianalisis, langkah selanjutnya adalah menentukan tujuan yang ingin dicapai dalam penelitian ini. Dalam tujuan ini, target yang ingin dicapai, terutama yang dapat mengatasi permasalahan yang ada. Adapun tujuannya adalah melakukan prediksi dengan probabilitas pada otorisasi sistem pengguna pada sistem berkas terkait adanya anomali data penggunaan.

Tinjauan Pustaka.
Kajian ini dilakukan untuk melengkapi perbendaharaan konsep, teori yang mendukung dalam pemecahannya masalah terutama kajian data mining dengan menggunakan algoritma klasifikasi pada teknik Naïve Bayes. Penelitian juga dilakukan melalui jurnal yang memiliki hubungan dengan penelitian dan referensi lainnya. Penelitian ini bertujuan untuk mengumpulkan data, baik data pokok maupun data pendukung, dimana semua data yang dibutuhkan dalam penelitian.

2.2. Metode Yang Diusulkan

Mengumpulkan Data.
Data yang dikumpulkan pada penelitian ini adalah data pelatihan (*data training*) dan data uji (*Data Testing*) pada konsep analisis Naïve Bayes. Dapat dijelaskan, semakin banyak data yang dikumpulkan, maka semakin baik dapat menyelesaikannya masalah. Adapun data yang dikumpulkan adalah dataset yang ada pada log aktifitas penggunaan sistem berkas oleh pengguna.

Analisis Data dengan Naïve Bayes
Langkah selanjutnya, data yang sudah terkumpul akan dilakukan proses analisa dengan menggunakan algoritma klasifikasi dengan teknik Naïve Bayes. Setiap data akan diberi dua bagian yaitu data pelatihan dan data uji. Setiap data akan diberikan analisis probabilitasnya.



Gambar 2. Tahapan Analisis Naïve Bayes

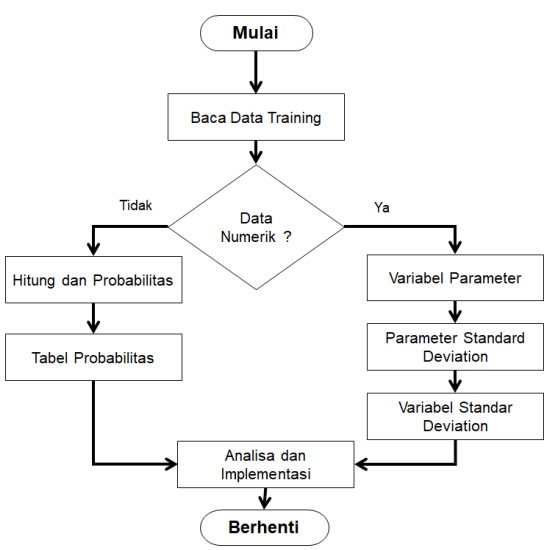
Pengolahan Data
Data yang diperoleh akan digunakan untuk analisis data pada metode Naïve Bayes. Setelah data selanjutnya analisis data yang dikumpulkan untuk menyesuaikan proses data dengan metode Naïve Bayes.

2.3. Implementasi dan Uji Coba

Data training dan data testing akan diproses dengan menggunakan konsep validasi yaitu cross validation. Adapun dalam tahap cross validasi digunakan 10 langkah pada data training dan data testing. Sedangkan pengujian akan dilakukan dengan

menggunakan confusion Matrix untuk memperoleh keakuratan model yang sudah diuji kevalidannya.

Gambar 3 menjelaskan skema proses analisa cros validation dengan Naïve Bayes. Pertama adalah membaca data training yang berisi data log aktifitas pada sistem file. Semua kegiatan pengguna direkam kemudian dianalisa. Tahap data training memerlukan waktu yang cukup lama sehingga data training memiliki beban yang cukup banyak. Selanjutnya, data training akan dibaca apakah bersifat numerik. Jika bersifat numerik maka akan dipersiapkan parameter yang terdapat pada variabel observasi. Selajutnya ditentukan parameter dari standar deviasi. Jika sebaliknya, data training yang bersifat selain numerik maka akan dihitung dan ditentukan nilai probabilitasnya selanjutnya dianalisa dari setiap nilai probabilitasnya.



Gambar 3. Skema Proses Cross Validation

Naïve Bayes Classifier merupakan sebuah metoda yang memiliki klasifikasi berakar pada teorema yang diusulkan oleh Bayes . Metode dengan kinerja pengklasifikasian menerapkan metode probabilitas dan statistik disajikan Thomas Bayes. Teori ini dapat memprediksi kehadiran peluang di masa depan berdasarkan pengalaman yang dibuktikan dengan data di masa sebelumnya sehingga dikenal sebagai Teorema Bayes. Karakteristik utama dari teorema probabilitas Naïve Bayes sebagai pengklasifikasi ini adalah asumsi yg sangat kuat akan independensi dari masing-masing kondisi atau kejadian.

Kegunaan yang menjadi utama adalah mengklasifikasikan dokumen berbentuk tidak terstruktur seperti teks berita ataupun teks bersifat akademis lainnya. Keunggulan kedua adalah sebagai yang tidak hanya pada metode machine learning yang menggunakan probabilitas namun diadopsi pada konsep data mining.

III. HASIL DAN PEMBAHASAN

Pada sistem berkas terdapat otorisasi yang diberikan kepada pengguna yang memungkinkan untuk mengakses sumber daya tertentu. Adapun bentuk sumber daya berada pada di jaringan, seperti file data, aplikasi, printer, dan pemindai. Izin pengguna juga menentukan jenis akses misalnya, apakah data hanya dapat dilihat (read only) atau dapat diperbarui (read/write).

Selanjutnya, dapat diketahui bahwa terdapat tipe izin yang diberikan pada sistem berkas yaitu (1) kontrol penu mendeskripsikan pengguna dapat membuka, menyimpan, memodifikasi, dan menghilangkan file dan subfolder. Selain itu, pengguna dapat mengubah pengaturan izin untuk semua file dan subdirektori. (2) modifikasi yaitu pengguna untuk membuka dan menyimpan file dan subfolder bahkan dapat kemungkinan penghapusan folder. (3) Membuka (Read) yaitu menjelaskan pengguna untuk melihat

dan menjalankan file yang dapat dieksekusi, termasuk skrip. (4) Membuka (Read) merupakan pengguna untuk melihat isi folder dan subfolder. (5) Menyimpan (Write) menjelaskan pengguna untuk menambahkan file dan subfolder.

3.1. Cross Validation

Pada Cross-validation dapat dianggap sebagai metode dalam pengolahan data statistik yang dapat diterapkan dalam mengevaluasi kinerja model atau algoritma. Selanjutnya, data dipisahkan menjadi dua subset yaitu data proses latihan dan data validasi yang disebut sebagai data testing. Model atau algoritma dilatih oleh subset latihan dan divalidasi oleh subset validasi. Selanjutnya pemilihan jenis cross validation dapat tentukan pada ukuran dataset. Pada sebagian peneliti mengungkapkan teknik Cross validation digunakan karena dapat mengurangi waktu komputasi dengan tetap menjaga keakuratan estimasi.

Dalam pembahasan berikutnya akan disajikan hasil validasi data trainning dan data testing. Dalam penelitian ini terdapat 162 data training dan 18 data testing. Dalam pemrosesannya terdapat 2 klasifikasi yaitu klasifikasi benar dan klasifikasi salah. Dengan demikian dari keduanya akan ditunjukan nilai akurasi seperti pada Tabel 1.

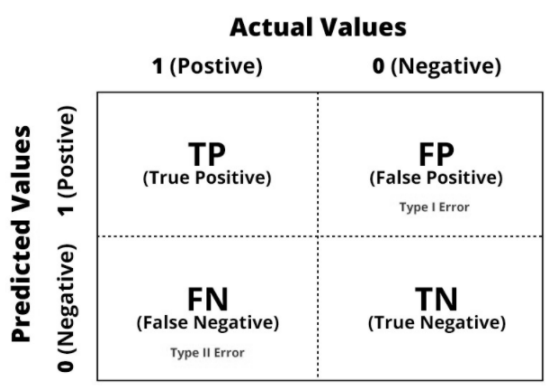
Tabel 1. Hasil Uji Cross Validation

Pengujian	Data Training	Data Testing	Klasifikasi Benar	Klasifikasi Salah
Uji 1	162	18	14	4
Uji 2	162	18	13	5
Uji 3	162	18	16	2
Uji 4	162	18	13	5
Uji 5	162	18	14	4
Uji 6	162	18	13	5
Uji 7	162	18	12	6
Uji 8	162	18	14	4
Uji 9	162	18	15	3
Uji 10	162	18	16	2

Hasil validasi yang menerapkan 10 iterasi pengujian disajikan pada tabel diatas memberikan hasil bahwa terdapat 162 data training dan 18 data testing. Dengan demikian dapat diberikan dua nilai akurasi yaitu akurasi benar dan akurasi salah. Nilai akurasi tertinggi diberikan pada pengujian pengujian 3 dan pengujian 10 sebesar 89% dan sebaliknya yang terendah pada pengujian 7.

3.2. Akurasi Model (Confusion Matrix)

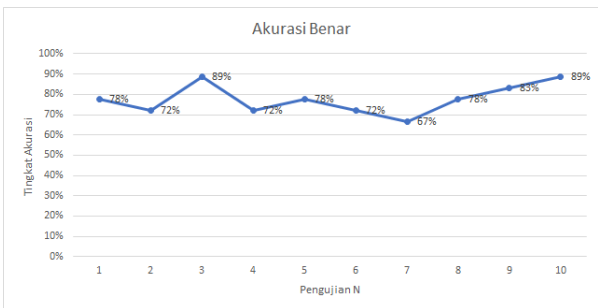
Terdapat 4 istilah sebagai representasi hasil proses klasifikasi pada confusion matrix. (1) True Positive (TP), (2) True Negative (TN), (3) False Positive, dan (4) (FP) False Negative (FN). Confusion Matrix merupakan salah satu pengukuran kinerja untuk masalah klasifikasi salah satunya machine learning. Bentuk luarannya dapat berupa dua kelas atau lebih. Confusion Matrix adalah tabel dengan 4 kombinasi berbeda dari nilai prediksi dan nilai aktual.



Gambar 4. Confusion Matrix

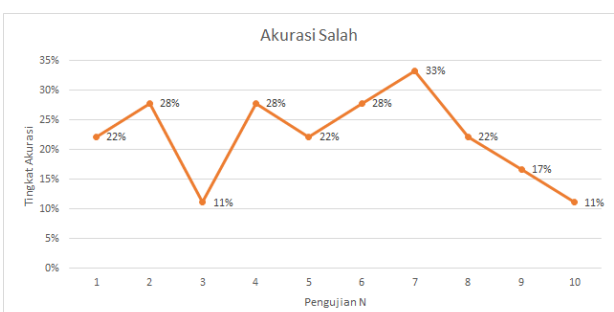
Dalam studi literatur lain disebutkan pada Confusion matrix sering dianggap sebagai error matrix. Pada hakikatnya metode confusion matrix mendeskripsikan tentang informasi komparasai hasil klasifikasi yang dilakukan oleh sistem (model) dengan hasil klasifikasi sebenarnya. Dalam konteks inni, confusion matrix berbentuk tabel matriks yang

menggambarkan kinerja model klasifikasi pada serangkaian data uji yang nilai sebenarnya diketahui. Gambar dibawah ini merupakan confusion matrix dengan 4 kombinasi nilai prediksi dan nilai aktual yang berbeda. Perhatikan gambar dibawah ini.



Gambar 5. Hasil Pengukuran Akurasi Benar

Hasil validasi yang menerapkan 10 iterasi pengujian disajikan pada tabel diatas memberikan hasil pada confusion matrix yang disajikan pada gambar diatas. Ada dua nilai akurasi yaitu akurasi benar dan akurasi salah. Nilai akurasi tertinggi diberikan pada pengujian pengujian 3 dan pengujian 10 sebesar 89% dan sebaliknya yang terendah pada pengujian 7.



Gambar 6. Hasil Pengukuran Akurasi Salah

Akurasi selanjutnya adalah akurasi dengan nilai yang salah dimulai dengan pengujian 1 dengan nilai 22% kemudian mengalami penurunan cukup signifikan mencapai 11% pada pengujian 3. Adapun nilai akurasi salah yang terbesar jatuh pada pengujian 7 mencapai 33% sehingga tetap tidak dianggap signifikan. Oleh sebab itu, dapat disimpulkan bahwa kinerja model dapat dianggap valid.

V. KESIMPULAN

Sistem keamanan file merupakan area yang kritis dan menjadi banyak target penyerangan dari pihak internal dan eksternal. Kemunculan anomali data berdasarkana log aktifitas penggunaan file dapat dijadikan sebagai analisis dan prediksi akan bahaya terhadap sistem berkas. Data mining dengan metode klasifikasi dan teknik analisis Naïve Bayes dapat dianggap cukup efektif terbukti dengan hasil validasi dan nilai akurasinya cukup tinggi. Oleh sebab itu, data mining menjadi sebuah area yang bisa dikembangkan dalam lingkup keamanan siber khususnya keamanan sistem berkas.

DAFTAR PUSTAKA

Annur, H. (2018). Klasifikasi Masyarakat Miskin Menggunakan Metode Naive Bayes. *Ilk. J. Ilm.*, 10(2), 160–165, 2018, doi: 10.33096/ilkom.v10i2.303.

Aridas, C. K., Karlos, S., Kanas, V. G. Fazakis, N. and Kotsiantis, S. B. (2020). Uncertainty Based Under-Sampling for Learning Naive Bayes Classifiers under Imbalanced Data Sets. *IEEE Access*,. 8, 2122–2133, doi: 10.1109/ACCESS. 2019.2961784.

Cai, N. H., Meng, Ryder, B. and Yao, D. (2019). DroidCat: Effective android

malware detection and categorization via app-level profiling. *IEEE Trans. Inf. Forensics Secur.*, 14(6), 1455–1470, 2019, doi: 10.1109/TIFS.2018.2879302.

Chazar C. and Widhiaputra, B. E. (2020). Machine Learning Diagnosa Kanker Payudara menggunakan Algoritma Support Vector Machine. *INFORMASI Jurnal Informatika dan Sistem Informasi, 12*(1), 67–78.

Das, A. K., Wazid, M., Kumar, N., Vasilakos, A. V., and Rodrigues, J. J. P. C. (2018). Biometrics-Based Privacy-Preserving User Authentication Scheme for Cloud-Based Industrial Internet of Things Deployment. *IEEE Internet Things J.*, 5(6), 4900–4913, doi: 10.1109/JIOT.2018.2877690.

Degirmenci, K. (2020). Mobile users' information privacy concerns and the role of app permission requests. *Int. J. Inf. Manage.*, 50, 261–272, doi: 10.1016/j.ijinfomgt.2019.05.010.

Dulhare, U. N. (2018). Prediction system for heart disease using Naive Bayes and particle swarm optimization. *Biomed. Res.*, 29(12), 2646–2649, doi: 10.4066/biomedicalresearch.29-18-620.

Fadlan, C., Ningsih, S. and Windarto, A. P. (2018). Penerapan Metode Naïve Bayes Dalam Klasifikasi Kelayakan Keluarga Penerima Beras Rastra. *J. Tek. Inform. Musirawas*, 3(1), p. 1, doi: 10.32767/jutim.v3i1.286.

Fadlil, A. Riadi, I. and Aji, S. (2017). DDoS Attacks Classification using Numeric Attribute-based Gaussian Naive Bayes. *Int. J. Adv. Comput. Sci. Appl.*, 8,(8), 42–50, doi: 10.14569/ijacsa.2017.080806.

Gunawan, (2019). Sistem Pendukung Keputusan Memilih Jurusan di Perguruan Tinggi Menggunakan Metode Analytical Hierarchy Process (AHP). *Inf. (Jurnal Inform. dan Sist. Informasi)*, 11(1), 1–17, doi: 10.37424/informasi.v11i1.7.

Gunduz M. Z. and Das, R. (2020). Cyber-security on smart grid: Threats and potential solutions. *Comput. Networks*, 169, p. 107094,, doi: 10.1016/j.comnet.2019.107094.

Imandasari, T., Irawan, E., Windarto, A. P. and Wanto, A. (2019). Algoritma Naive Bayes Dalam Klasifikasi Lokasi Pembangunan Sumber Air. *Pros. Semin. Nas. Ris. Inf. Sci.*, 1, p. 750, 2019, doi: 10.30645/senaris.v1i0.81.

Irmayani, W. (2021). Visualisasi Data Pada Data Mining Menggunakan Metode Klasifikasi. *J. KHATULISTIWA Inform.*, IX(I), 68–72.

Laksono, R. A., Sungkono, K. R., Sarno, R. and Wahyuni, C. S. (2019). Sentiment analysis of restaurant customer reviews on tripadvisor using naïve bayes. *Proc. 2019 Int. Conf. Inf. Commun. Technol. Syst. ICTS 2019*, 49–54, doi: 10.1109/ICTS.2019.8850982.

Mustafa, M. S., Ramadhan, M. R. and Thenata, A. P. (2017). Implementasi Data Mining untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan

Algoritma Naive Bayes Classifier. *Citec J.*, 4(2), 151–162.

Parlina I., Arnol, M.Y., Febriati, N.A., Dewi, R., Wanto, A., Lubis, M.R., Susiani. (2019). Naive Bayes Algorithm Analysis to Determine the Percentage Level of visitors the Most Dominant Zoo Visit by Age Category. *J. Phys. Conf. Ser.*, 1255, 012031, doi: 10.1088/1742-6596/1255/1/012031.

Peling, I. B. A. Arnawan, I. N. Arthawan, I. P. A. and Janardana, I. G. N. (2017). Implementation of Data Mining To Predict Period of Students Study Using Naive Bayes Algorithm. *Int. J. Eng. Emerg. Technol.*, 2(1), p. 53, doi: 10.24843/ijeet.2017.v02.i01.p11.

Pranckevičius, T. and Marcinkevičius, V. (2017). Comparison of Naive Bayes, Random Forest, Decision Tree, Support Vector Machines, and Logistic Regression Classifiers for Text Reviews Classification," *Balt. J. Mod. Comput.*, 5(2), 221–232, doi: 10.22364/bjmc.2017.5.2.05.

Putri, S.U., Irawan, E. & Rizky, F. (2021). Implementasi Data Mining Untuk Prediksi Penyakit Diabetes. *KESATRIA (J. penerapan Sist. Inf. dan Manaj.*, 2(1), 39–46.

Rasjid Z. E. and Setiawan, R. (2017). Performance Comparison and Optimization of Text Document Classification using k-NN and Naïve Bayes Classification Techniques," *Procedia Comput. Sci.*, 116, 107–112, doi: 10.1016/j.procs.2017.10.017.

Rifqo, M. H. and Wijaya, A. (2017). Implementasi Algoritma Naive Bayes Dalam Penentuan Pemberian Kredit. *Pseudocode*, 4(2), 120–128, doi: 10.33369/pseudocode.4.2.120-128.

Slamet, C., Andrian, R., Maylawati, D. S. Suhendar, Darmalaksana, W., and Ramdhani, M. A. (2018). Web Scraping and Naïve Bayes Classification for Job Search Engine. *IOP Conf. Ser. Mater. Sci. Eng.*, 288, doi: 10.1088/1757-899X/288/1/012038.Supriyatna A. and Mustika, W. P. (2018). Komparasi Algoritma Naive bayes dan SVM Untuk Memprediksi Keberhasilan Imunoterapi Pada Penyakit Kutil. *J-SAKTI (Jurnal Sains Komput. dan Inform.*, 2(2) p. 152, , doi: 10.30645/j-sakti.v2i2.78.

Taylor, P. J., Dargahi, T., Dehghantanha, A., Parizi, R. M. and Choo, K. K. R. (2020) A systematic literature review of blockchain cyber security *Digit. Commun. Networks*, 6(2), 147–156, doi: 10.1016/j.dcan.2019.01.005.

Wahyuningsih S. and Utari, D. R. (2018). Perbandingan Metode K-Nearest Neighbor, Naive Bayes dan Decision Tree untuk Prediksi Kelayakan Pemberian Kredit. *Konf. Nas. Sist. Inf. 2018 STMIK Atma Luhur Pangkalpinang, 8 – 9 Maret 2018*, pp. 619–623.

Wongkar, M. and Angdresey, A. (2019). Sentiment Analysis Using Naive Bayes Algorithm Of The Data Crawler: Twitter. *Proc. 2019 4th Int. Conf. Informatics Comput. ICIC 2019*. 1–5, doi: 10.1109/ICIC47613.2019.8985884.

Wood, A., Shpilrain, V., Najarian, K. and Kahrobaei, D. (2019). Private naive bayes classification of personal biomedical data: Application in cancer data Application in cancer data analysis," *Comput. Biol. Med.*, 105, 144–150, doi: 10.1016/j.compbiomed.2018.11.018.