JURNAL INSTEK
INFORMATIKA SAINS DAN TEKNOLOGI
Volume 9 Nomor. 1 April 2024
P –ISSN : 2541-1179, E-ISSN : 2581-1711
Ojs :http://journal.uin-alauddin.ac.id/index.php/instek/index
Email : instek@uin-alauddin.ac.id

ANALISIS KOMPARASI PEMODELAN KLASIFIKASI PADA DATASET PENERIMAAN MAHASISWA BARU PERGURUAN TINGGI

MUHAMMAD WAHYUDI[1], EDI SURYA NEGARA[2]
[1,2] Program Studi Magister Teknik Informatika,
Fakultas Teknik, Universitas Bina Darma
Email: [1]wahyudiabuhasan@gmail.com, [2]e.s.negara@binadarma.ac.id

ABSTRAK

Data Penerimaan Mahasiswa Baru (PMB) merupakan aset yang memiliki nilai kebermanfaatan bagi pemangku kepentingan di Perguruan Tinggi. Ekstraksi pengetahuan dari dataset PMB dapat memberikan rekomendasi strategi sesuai dengan kondisi dan fakta yang ada. Pada penelitian ini dilakukan pemodelan klasifikasi pada dataset PMB Perguruan Tinggi untuk melihat preferensi pemilihan program studi bagi mahasiswa baru. Tahapan yang dilakukan meliputi *Business Understanding*, *Data Understanding*, *Data Pre-Processing*, *Modeling* dan Evaluasi. Dataset yang diolah sejumlah 2704 record data dengan melibatkan 5 model yang dikomparasi, yakni kNN, *Naïve Bayes Classifier, Decision Tree Classifier, Support Vector Machine dan AdaBoost*. Evaluasi menggunakan nilai Akurasi dan F1 Score. Hasil akhir menunjukkan tingkat akurasi sebesar 74% pada *Decision Tree Classifier*, *Support Vector Machine dan Adaptive Boosting (AdaBoost).*

Kata Kunci: Pemodelan, Klasifikasi, Dataset PMB

I. PENDAHULUAN

Penerimaan Mahasiswa Baru (PMB) di suatu Perguruan Tinggi merupakan tahapan awal dalam menghasilkan Sumber Daya Manusia yang unggul dan kompetitif. Strategi dalam meningkatkan layanan PMB dilakukan secara berkelanjutan guna mencapai tata kelola Perguruan Tinggi yang *"Good Governance"* (Ermini & Nindiati, 2019). Beberapa upaya strategi yang telah diimplementasikan diantaranya adalah digitalisasi proses PMB dengan adanya sistem penerimaan mahasiswa baru yang terintegrasi dengan sistem komputer, rekayasa agenda penerimaan mahasiswa baru, dan strategi pemasaran yang tepat sasaran (Djogo, 2020).

Dengan banyaknya aktifitas yang dilakukan oleh PMB melalui strategi yang diimplementasikan, tentunya menghasilkan data dari calon mahasiswa, diantaranya adalah data pendaftar calon mahasiswa baru. Umumnya data tersebut digunakan pada saat proses pendaftaran dan menjadi informasi awal bagian akademik untuk menetapkan calon mahasiswa menjadi mahasiswa aktif perguruan

JURNAL INSTEK
INFORMATIKA SAINS DAN TEKNOLOGI
Volume 9 Nomor. 1 April 2024
P –ISSN : 2541-1179, E-ISSN : 2581-1711
Ojs :http://journal.uin-alauddin.ac.id/index.php/instek/index
Email : instek@uin-alauddin.ac.id

tinggi. Setelahnya, seluruh data dan dokumen PMB sebatas menjadi arsip dan bagian dari proses pelaporan dari pihak divisi PMB yang pada akhirnya menjadi data historis yang belum memiliki nilai manfaat yang signifikan. Salah satu upaya dalam pemanfaatan data historis tersebut adalah dengan mengimplementasikan *data science* melalui *Machine Learning* yang dapat mengeksplorasi data lebih dalam.

Data historis PMB sangat berpotensi untuk dianalisis lebih dalam melalui *machine learning* guna mengetahui informasi yang tersirat dari serangkaian dataset. Salah satu metode yang digunakan adalah klasifikasi (Amanda & Negara, 2020), dimana dalam metode klasifikasi dilakukan pembagian 2 elemen, yakni fitur *predictor* yang akan memprediksi kelas berdasarkan label tertentu (Leidiyana & Permana, 2020). Metode klasifikasi termasuk ke dalam *supervised learning* karena terdapat label yang mendeskripsikan fitur-fiturnya (Qisthiano et al., 2021). Beberapa metode klasifikasi yang umumnya digunakan dalam pemodelan diantaranya adalah *Decision Tree Classifier, Naïve Bayes Classifier, Support Vector Machine* (Hidayat & Yulianingsih, 2021).

Dataset didalam penelitian ini merupakan data historis PMB Perguruan Tinggi di Kota Sukabumi dengan jumlah data *record* sebanyak 2704 data untuk periode 2019, 2020 dan 2021. Terdapat lebih dari 20 fitur di dalam dataset yang terbagi ke dalam 2 kategori utama, yakni: (1) Data diri pendaftar yang terdiri dari Nama, Jenis Kelamin, Agama, Tempat Lahir, Tanggal Lahir, Status Sipil, Alamat, Kodepos, Provinsi, Kota, dan seterusnya; (2) Latar belakang pendidikan pelamar diantaranya Jenis Sekolah, Nama Sekolah Jurusan, Jurusan Sekolah, Nilai Unas, Tanggal Lulus, Tahun Lulus, dan seterusnya. Dari fitur tersebut dapat dicermati bahwa label kelas dalam penelitian ini adalah Program Studi. Namun dari banyak nya fitur atau *predictor* yang ada, perlu dicermati yang manakah memiliki nilai keberpengaruhan tinggi yang mendukung dalam klasifikasi Program Studi tersebut.

Dalam aspek metode klasifikasi yang diimplementasikan, perlu dikaji lebih dalam bahwa model manakah yang memiliki nilai akurasi terbaik untuk merepresentasikan hasil yang valid dan representatif dalam konteks dataset yang

JURNAL INSTEK
INFORMATIKA SAINS DAN TEKNOLOGI
Volume 9 Nomor. 1 April 2024
P –ISSN : 2541-1179, E-ISSN : 2581-1711
Ojs :http://journal.uin-alauddin.ac.id/index.php/instek/index
Email : instek@uin-alauddin.ac.id

dikaji yakni dataset PMB periode 2019, 2020 dan 2021. Sehingga disimpulkan bahwa rumusan penelitian adalah (1) Bagaimana fitur *predictor* yang memiliki nilai keterhubugan terbaik untuk diimplementasikan dalam pemodelan klasifikasi program studi?, dan; (2) Bagaimana pemodelan klasifikasi dengan akurasi terbaik yang diimplementasikan dalam klasifikasi program studi?. Sehingga tujuan dari penelitian ini adalah (1) Menentukan fitur *predictor* dalam pengklasifikasian program studi pada dataset PMB Perguruan Tinggi Kota Sukabumi; (2) Menentukan rekomendasi pemodelan klasifikasi yang diimplementasikan dalam klasifikasi program studi dataset PMB Perguruan Tinggi Kota Sukabumi.

II. METODE PENELITIAN

Penelitian yang dilakukan menggunakan pendekatan data life cycle CRISP DM yang terdiri dari tahapan *Business Understanding*, *Data Understanding*, *Data Preprocessing*, *Modeling* dan *Evaluasi.* Objek penelitian merupakan dataset PMB Perguruan Tinggi dengan jumlah sebanyak 2704 data selama tahun 2019, 2020 dan 2021. Metode klasifikasi yang dikomparasi terdiri dari k-Nearest Neighbor (k-NN), Naïve Bayes Classifier, Decision Tree Classifier, Support Vector Machine dan Adaptive Boosting (AdaBoost). Berikut adalah penjelasan teknis dari tahapan penelitian yang dilakukan:

A.*Business Understanding*

Di dalam *Business Understanding* dilakukan tahapan mendasar dalam memahami konteks permasalahan yang akan dicermati, tujuan, solusi dari perspektif bisnis dan instrumen pengukuran keberhasilan yang digunakan dalam menentukan kriteria pemodelan yang optimum (Gunawan, 2021). Dalam kasus ini permasalahan yang diangkat adalah bagaimana mengoptimalkan dataset PMB Perguruan Tinggi Kota Sukabumi menjadi aset yang memiliki daya guna, dimana dalam konteks ini adalah mengklasifikasikan program studi berdasarkan fitur prediktor dalam dataset PMB. Tujuan dari penelitian merupakan dasar dalam menentukan target akhir apa yang ingin dicapai, yakni optimalisasi dataset PMB Perguruan Tinggi Kota Sukabumi melalui pemodelan klasifikasi yang optimal dalam pengklasifikasian program studi.

JURNAL INSTEK
INFORMATIKA SAINS DAN TEKNOLOGI
Volume 9 Nomor. 1 April 2024
P –ISSN : 2541-1179, E-ISSN : 2581-1711
Ojs :http://journal.uin-alauddin.ac.id/index.php/instek/index
Email : instek@uin-alauddin.ac.id

B.*Data Understanding*

Tahapan *data understanding* merupakan proses penelaahan data, dimana dilakukan *Exploration Data Analysis (EDA)* (Isa & Junedi, 2022). Tahapan ini bertujuan untuk melihat anomali yang terdapat dalam dataset, seperti *missing data*, redundansi data, *outlier*, serta *imbalance* data (Muhammad, 2019). Identifikasi anomali data dalam dataset dapat memberikan gambaran bagaimana kondisi dataset secara kompleks, untuk meminimalisir bahkan menghilangkan *noise* yang terdapat dalam dataset PMB. Sehingga dataset PMB tersebut menghasilkan tingkat akurasi dan kredibilitas yang tinggi. Berikut fitur yang dicermati dalam dataset PMB yang terbagi ke dalam 2 besar kategori, yakni data diri pendaftar dan Pendidikan.

C.*Data Pre-Processing*

Setelah dilakukan penelaahan data, selanjutnya adalah tahapan *data pre-processing* dimana dilakukan transformasi data berdasarkan temuan-temuan yang dihasilkan pada tahapan *data understanding*. Proses transformasi dilakukan dengan menggunakan beberapa teknik, yang tergantung dari jumlah data anomali. Beberapa diantaranya adalah dengan melakukan penghapusan data *(delete record)*, proses imputasi data dengan mengisi *missing value*, *balancing* data dengan menambah sebaran data sehingga *record* pada fitur menjadi berimbang (Suad A. & Wesam S., 2017).

D.*Modeling*

Metode yang digunakan dalam *modeling* adalah klasifikasi, dimana akan mengkomparasi beberapa model klasifikasi, yakni k-Nearest Neighbor (k-NN), *Naïve Bayes Classifier, Decision Tree Classifier, Support Vector Machine (SVM) dan Adaptive Boosting (AdaBoost).* Secara teori kNN dilakukan dengan menghitung jarak terdekat antara objek dan dilakukan pengklasifikasian terhadap objek-objek tersebut (Anam et al., 2021). Naïve Bayes Classifier menggunakan pendekatan probabilitas dalam melakukan proses pengklasifikasian (Isa, 2021), sedangkan Decision Tree Classifier menggunakan pemodelan pohon dalam menentukan klasifikasi (Singh & Gupta, 2019). SVM dan AdaBoost merupakan mengembangan lebih lanjut dalam metode klasifikasi, dimana SVM menggunakan

JURNAL INSTEK
INFORMATIKA SAINS DAN TEKNOLOGI
Volume 9 Nomor. 1 April 2024
P –ISSN : 2541-1179, E-ISSN : 2581-1711
Ojs :http://journal.uin-alauddin.ac.id/index.php/instek/index
Email : instek@uin-alauddin.ac.id

fungsi linear dalam fitur yang berdimensi tinggi (Rachmatika & Bisri, 2020).

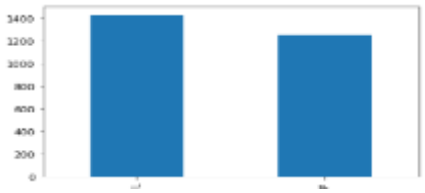
E.Evaluasi

Evaluasi yang dilakukan dalam penelitian ini menggunakan *confusion matrix* untuk melihat kinerja dari model yang diimplementasikan, terdapat 4 aspek dalam *confusion matrix* dalam kondisi prediktif dan actual yakni *True Positive*, *True Negative*, *False Positive* dan *False Negative* (Hossin & Sulaiman, 2015). Dari perhitungan *confusion matrix* dapat dilihat 4 penilaian kinerja, yakni *accuracy, precision, recall* dan F1 *Score*.

III. HASIL DAN PEMBAHASAN

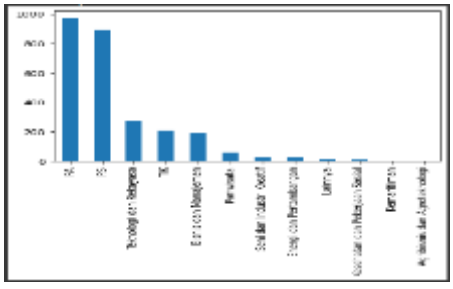
A.*Exploratory Data Analysis (EDA)*

Dalam penelaahan data dilakukan *Exploratory Data Analysis* (EDA) yang bertujuan untuk mencermati fitur manakah yang memiliki korelasi signifikan terhadap kelas Program Studi. Misalnya pada fitur Jenis Kelamin dapat dilihat grafik jumlah Laki-laki dan perempuan pada gambar 1 berikut:



Gambar 1. Fitur Jenis Kelamin pada Dataset PMB

Analisis fitur lainnya, misalnya pada "Jurusan Sekolah" menghasilkan 12 data label yang terdiri dari IPA, IPS, Teknologi dan Rekayasa, TIK, Bisnis dan Manajemen, Pariwisata, Seni dan Industri Kreatif, Energi dan Pertambangan, Lainnya, Kesehatan dan Pekerjaan Sosial, Kemaritiman, dan Agribisnis dan Agroteknoloti. Keseluruhan data tersebut jika divisualisasikan, ditunjukkan pada gambar 2 berikut:



Gambar 2. Visualisasi "Jurusan Sekolah"

JURNAL INSTEK
INFORMATIKA SAINS DAN TEKNOLOGI
Volume 9 Nomor. 1 April 2024
P –ISSN : 2541-1179, E-ISSN : 2581-1711
Ojs :http://journal.uin-alauddin.ac.id/index.php/instek/index
Email : instek@uin-alauddin.ac.id

Dari visualisasi gambar 2 tersebut dapat dilihat bahwa 2 jurusan yang mendominasi adalah IPA dan IPS dengan rentang data lebih dari 800, jurusan teknologi dan rekayasa dengan data lebih dari 200, sedangkan jurusan lainnya hanya memiliki data di bawah 200. Jika dicermati lebih lanjut, guna memberikan efektifitas dalam akurasi pemodelan, jurusan diluar IPA dan IPS dapat digabungkan menjadi "Lainnya" sehingga tidak terjadi gap data yang terlalu jauh antara IPA/IPS dengan jurusan lainnya. Sehingga pada fitur "Jurusan Sekolah" terbagi menjadi 3 kategori utama yakni IPA, IPS, dan Lainnya. Secara ringkas hasil analisis fitur pada dataset PMB Perguruan Tinggi Kota Sukabumi yang direkomendasikan untuk pemodelan berdasarkan EDA dapat dilihat pada tabel 1 berikut:

Tabel 1. Hasil *Exploratory Data Analysis* pada Dataset PMB

No	Nama Fitur	Temuan	Rekomendasi Fitur
1	Jenis Kelamin	Sebaran hampir berimbang dengan jumlah L = 1437 dan P=1266	Dapat Direkomendasikan
2	Penghasilan	Data kategori dengan terdiri dari 5 kategori	Dapat Direkomendasikan, perlu *balancing data* dan mempersempit kategori
3	Jurusan Sekolah		Dapat Direkomendasikan, perlu dikategorikan lebih general
4	Jenis Sekolah	Data terdiri dari SMA dan SMK yang dapat dikategorikan	Dapat Direkomendasikan. Fitur dapat dipecah menjadi Tipe Sekolah (SMA/SMK/Lainnya) dan Status Sekolah (Negeri/ Swasta/ Lainnya)
5	Program Studi	Target Label	Target Label

Selanjutnya dilakukan analisis chi square untuk melihat korelasi antara fitur rekomendasi dengan target label dimana dalam hal ini adalah "Program Studi" sebagai variabel dependen dengan variable independen yakni "Jenis Kelamin",

JURNAL INSTEK
INFORMATIKA SAINS DAN TEKNOLOGI
Volume 9 Nomor. 1 April 2024
P –ISSN : 2541-1179, E-ISSN : 2581-1711
Ojs :http://journal.uin-alauddin.ac.id/index.php/instek/index
Email : instek@uin-alauddin.ac.id

"Penghasilan", "Jurusan Sekolah", dan "Jenis Sekolah". Hasil dari analisis chi square dari masing-masing fitur parameter terhadap fitur label (Program Studi) ditunjukkan pada tabel 2 berikut:

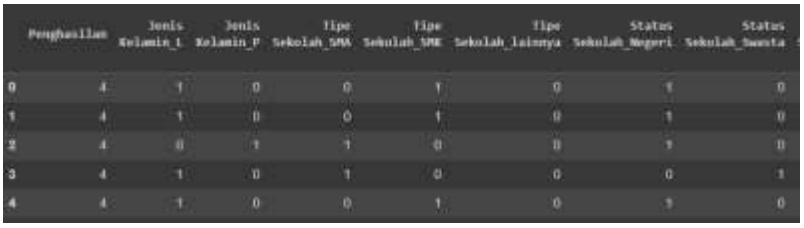
Tabel 2. Hasil pengujian chi square

No	Fitur	Chi Skor
1	Jenis Kelamin	2.90
2	Penghasilan	1.03
3	Jurusan Sekolah	3.91
4	Jenis Sekolah	1.95

Secara berurutan, peringkat tertinggi chi skor adalah pada fitur "Jurusan Sekolah" dengan chi skor sebesar 3.91. Peringkat kedua adalah fitur "Jenis Kelamin dengan skor 2.90. Sedangkan yang terendah adalah fitur "Penghasilan" dengan skor sebesar 1.03.

B.Hasil Pemodelan

Dalam proses pemodelan dilakukan *one hot encoding* sebagai proses transformasi data dari data teks menjadi data *binary* dimana melibatkan fitur yang direkomendasikan pada tahapan EDA sebelumnya, yakni "Penghasilan", "Jenis Kelamin", "Tipe Sekolah", "Status Sekolah", dan "Bidang Jurusan". Gambar 4 menunjukkan fitur setelah dilakukan proses *one hot encoding*:



Gambar 4. Fitur setelah *One Hot Encoding*

Berikutnya implementasi pemodelan dari dataset mahasiswa baru dilakukan dengan menentukan kriteria dari data yang akan dimodelkan dengan persentasi data latih sebanyak 80% dan data uji sebesar 20%. Dari hasil pemodelan yang diimplementasikan didapatkan hasil penilaian berupa akurasi dan F1 Score dengan nilai tertinggi akurasi pada model Decision Tree Classifier dan Support Vector Machine sebesar 25%, dan hasil *F1 Score* dengan nilai tertinggi sebesar 16% untuk model Naïve Bayes Classifier dan Decision Tree Classifier sebagaimana

JURNAL INSTEK
INFORMATIKA SAINS DAN TEKNOLOGI
Volume 9 Nomor. 1 April 2024
P –ISSN : 2541-1179, E-ISSN : 2581-1711
Ojs :http://journal.uin-alauddin.ac.id/index.php/instek/index
Email : instek@uin-alauddin.ac.id

ditunjukkan pada tabel 3 berikut:

Tabel 3. Hasil Pemodelan

No	Model	Akurasi	***F1 Score***
1	k-Nearest Neighbor (k-NN)	18%	17%
2	Naïve Bayes Classifier	24%	16%
3	Decision Tree Classifier	25%	16%
4	Support Vector Machine (SVM)	25%	15%
5	Adaptive Boosting (AdaBoost)	20%	13%

Hasil pengujian model tersebut tidak menunjukan hasil yang signifikan, jika dianalisis lebih lanjut hal ini disebabkan karena sebaran kelas label yang beragam dimana terdapat 14 label program studi yang tidak diimbangi dengan jumlah dataset yang memadai. Sehingga diperlukan rekonstruksi label atau *re-labeling* dengan mengklasifikasikan kelas label program studi dibagi ke dalam 2 label klasifikasi, yakni "Ilmu Teknik dan Terapan" dan "Ilmu Sosial". Ilmu Teknik dan Terapan terdiri dari Teknik Sipil, Teknik Informatika, Survei dan Pemetaan, Sistem Komputer, Sistem Informasi, Perencanaan Wilayah dan Kota, Manajemen Informatika, Desain Komunikasi Visual, Arsitektur, Keselamatan dan Kesehatan Kerja. Sedangkan Ilmu Sosial terdiri dari Akuntansi, Manajemen, Pendidikan Bahasa Inggris, dan Ilmu Pemerintahan. Selanjutnya adalah memodelkan kembali dari dataset yang sudah dilakukan *re-labeling* pada kelas program studi tersebut. Hasil dari pemodelan menunjukkan peningkatan akurasi dan F1 *Score* dengan nilai akurasi tertinggi sebesar 74% pada Decision Tree Classifier, Support Vector Machine, dan Adaptive Boosting (AdaBoost). Sedangkan F1 *Score* tertinggi sebesar 71% pada Naïve Bayes Classifier. Secara lengkap dapat dilihat pada tabel 4 berikut:

Tabel 4. Hasil Pemodelan setelah *re-labeling* Kelas "Program Studi"

No	Model	Akurasi	***F1 Score***
1	k-Nearest Neighbor (k-NN)	69%	67%
2	Naïve Bayes Classifier	72%	71%
3	Decision Tree Classifier	74%	67%
4	Support Vector Machine (SVM)	74%	67%
5	Adaptive Boosting (AdaBoost)	74%	67%

IV. KESIMPULAN

Hasil akhir pemodelan menunjukan nilai akurasi yang signifikan yakni sebesar

JURNAL INSTEK
INFORMATIKA SAINS DAN TEKNOLOGI
Volume 9 Nomor. 1 April 2024
P –ISSN : 2541-1179, E-ISSN : 2581-1711
Ojs :http://journal.uin-alauddin.ac.id/index.php/instek/index
Email : instek@uin-alauddin.ac.id

74% untuk Decision Tree Classifier, Support Vector Machine dan Adaptive Boosting yang disebabkan pada ketiga model tersebut memiliki kompleksitas pemrosesan data yang lebih tinggi jika dikomparasi dengan model lainnya. *Relabeling* kelas Program Studi dilakukan dengan mengklasifikasi program studi menjadi 2 klasifikasi, yakni "Ilmu Teknik dan Terapan" dan "Ilmu Sosial". Akurasi di atas 70% menunjukkan model dapat digunakan dalam melakukan prediksi pemilihan program studi bagi mahasiswa baru dengan menggunakan fitur Penghasilan, Jenis Kelamin, Tipe Sekolah, Status Sekolah dan Bidang Jurusan. Sebagai saran untuk pengembangan berikutnya diperlukan *parameter tuning* pada setiap fitur sehingga meningkatkan kinerja akurasi dari model yang dibangun.

DAFTAR PUSTAKA

Amanda, R., & Negara, E. S. (2020). Analysis and Implementation Machine Learning for YouTube Data Classification by Comparing the Performance of Classification Algorithms. *Jurnal Online Informatika*, *5*(1), 61–72. https://doi.org/10.15575/join.v5i1.505

Anam, M. K., Pikir, B. N., & Firdaus, M. B. (2021). Penerapan Na ̈ive Bayes Classifier, K-Nearest Neighbor (KNN) dan Decision Tree untuk Menganalisis Sentimen pada Interaksi Netizen danPemeritah. *MATRIK : Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, *21*(1), 139–150. https://doi.org/10.30812/matrik.v21i1.1092

Djogo, Y. O. (2020). Strategi Marketing Penerimaan Mahasiswa Baru Universitas Sanggabuana Bandung di Tengah Pandemi Covid-19. *Business Preneur: Jurnal Ilmu Administrasi*, *2*(2), 88–100.

Ermini, E., & Nindiati, D. S. (2019). Pengaruh Kualitas Pelayanan Panitia Penerimaan Mahasiswa Baru terhadap Kepuasan Calon Mahasiswa di Universitas PGRI Palembang. *International Journal of Social Science and Business*, *3*(4), 532. https://doi.org/10.23887/ijssb.v3i4.21767

Gunawan, G. (2021). Data Mining Using Crisp-Dm Process Framework on Official Statistics: a Case Study of East Java Province. *Jurnal Ekonomi Dan Pembangunan*, *29*(2), 183–198. https://doi.org/10.14203/jep.29.2.2021.183-198

Hidayat, R., & Yulianingsih, E. (2021). Penerapan Data Mining untuk Identifikasi Penyakit DBD menggunakan Metode Klasifikasi (Studi Kasus: Rumah Sakit Tk II 02.05.01 dr. Ak Gani Palembang). *Jurnal Bina Komputer*, *3*(2), 9–11.

Hossin, M., & Sulaiman, M. . (2015). A Review on Evaluation Metrics for Data Classification Evaluations. *International Journal of Data Mining & Knowledge Management Process*, *5*(2), 01–11. https://doi.org/10.5121/ijdkp.2015.5201

Isa, I. G. T. (2021). Aplikasi Asesmen Calon Debitur menggunakan Naive Bayes di Koperasi Mitra Sejahtera SMK Negeri 1 Kota Sukabumi. *Jurnal Sisfokom*

JURNAL INSTEK
INFORMATIKA SAINS DAN TEKNOLOGI
Volume 9 Nomor. 1 April 2024
P –ISSN : 2541-1179, E-ISSN : 2581-1711
Ojs :http://journal.uin-alauddin.ac.id/index.php/instek/index
Email : instek@uin-alauddin.ac.id

(Sistem Informasi Dan Komputer), *10*(1), 31–39. https://doi.org/10.32736/sisfokom.v10i1.1013

Isa, I. G. T., & Junedi, B. (2022). Hyperparameter Tuning Epoch dalam Meningkatkan Akurasi Data Latih dan Data Validasi pada Citra Pengendara. *Seminar Nasional Sains Dan Teknologi2*, 231–237.

Leidiyana, H., & Permana, A. A. (2020). Pemodelan Klasifikasi Dalam Meningkatkan Proses Pemilihan Calon Karyawan dengan Metode C4.5 dan Jaringan Syaraf Tiruan. *Jurnal Teknik Informatika (JIKA) Universitas Muhammadiyah Tangerang*, *3*(2), 7–14.

Muhammad, B. (2019). Implementasi Data Mining untuk Prediksi Standar Hidup Layak Berdasarkan Tingkat Kesehatan dan Pendidikan Masyarakat. *Jurnal SISKOM-KB (Sistem Komputer Dan Kecerdasan Buatan)*, *2*(2), 33–37. https://jurnal.tau.ac.id/index.php/siskom-kb/article/view/58

Qisthiano, M. R., Kurniawan, T. B., Negara, E. S., & Akbar, M. (2021). Pengembangan Model Untuk Prediksi Tingkat Kelulusan Mahasiswa Tepat Waktu dengan Metode Naïve Bayes. *Jurnal Media Informatika Budidarma*, *5*(3), 987. https://doi.org/10.30865/mib.v5i3.3030

Rachmatika, R., & Bisri, A. (2020). Perbandingan Model Klasifikasi untuk Evaluasi Kinerja Akademik Mahasiswa. *Jurnal Edukasi Dan Penelitian Informatika (JEPIN)*, *6*(3), 417. https://doi.org/10.26418/jp.v6i3.43097

Singh, G. A. P., & Gupta, P. K. (2019). Performance analysis of various machine learning-based approaches for detection and classification of lung cancer in humans. *Neural Computing and Applications*, *31*(10), 6863–6877. https://doi.org/10.1007/s00521-018-3518-x

Suad A., A., & Wesam S., B. (2017). Review of data preprocessing techniques in data mining.pdf. *Journal of Engineering and Applied Sciences*, *12*(16), 4102–4107. https://doi.org/doi=jeasci.2017.4102.4107