

ANALISIS PREDIKSI KELULUSAN UJIAN LABOR PROFESIONAL MAHASISWA UNIVERSITAS METAMEDIA

KARFINDO¹, RIFA TURAINA², RUSLI SAPUTRA³

Sistem Informasi, Universitas Metamedia¹

karfindo@stmikindonesia.ac.id¹, rifaturaina@stmikindonesia.ac.id²,

ruslisaputra@stmikindonesia.ac.id¹

Abstract : *Increasing developments in the field of Artificial Intelligence (AI) have made many other fields begin to apply AI in data analysis in their respective fields, such as health, finance, education, and others. In the field of education, it is currently known as Education Data Mining (EDM) which is a scientific discipline for exploring data originating from educational contexts. At metamedia university there are already many information systems used to process student data. With this data warehouse, the authors try to perform data analysis using the CRIPS-DM method. CRIPS-DM is an industry independent process model for data mining. One of the problems that occurs is failure in the professional labor exam. The author tries to apply the naïve Bayes machine learning algorithm to analyze the causes of student failure. The data that is processed is data on the value of labor exams, data on the value of 6 courses. To evaluate the accuracy of the writer using a confusion matrix with an accuracy rate of 68.75%*

Keywords: *Artificial Intelligence, Education Data Mining, Naïve Bayes, professional laboratory exams.*

Abstrak: Peningkatan perkembangan dibidang *Artificial Intelegence* (AI), membuat banyak bidang lainnya mulai menerapkan AI dalam analisis data yang dimiliki pada masing-masing bidang, seperti bidang kesehatan, keuangan, pendidikan, dan lain-lain. Di dalam bidang pendidikan saat ini dikenal dengan *Ecuation Data Mining* (EDM) yang merupakan disiplin ilmu untuk mengeksplorasi data yang berasal dari konteks pendidikan. Di universitas metamedia sudah banyak sistem informasi yang digunakan untuk mengolah data-data mahasiswa. Dengan adanya gudang data tersebut maka penulis mencoba untuk melakukan analisis data dengan menggunakan metode CRIPS-DM. CRIPS-DM adalah model proses independen industri untuk data mining. Salah satu masalah yang terjadi adalah kegagalan dalam ujian labor profesional. Penulis mencoba menerapkan algoritma machine learning naïve bayes untuk melakukan analisis penyebab kegagalan mahasiswa. Data yang diolah adalah data-data nilai ujian labor, data nilai 6 matakuliah. Untuk melakukan evaluasi akurasi penulis menggunakan confusion matrix dengan tingkat akurasi sebesar 68.75 %.

Kata Kunci : *Artifical Intelegence, Education Data Mining, Naïve Bayes, ujian labor profesional.*

A.Pendahuluan

Artifical intelegence (AI) adalah program yang memiliki bentuk matematis (instruksi). Berbeda dengan program biasa yang menghasilkan aksi berdasarkan instruksi, tujuan kecerdasan buatan adalah menciptakan program yang mampu memprogram atau kemampuan untuk belajar. Dengan terjadinya peningkatan perkembangan dibidang AI, maka banyak bidang lainnya yang mulai menerapkan AI dalam analisis data yang dimiliki pada masing-masing bidang, seperti bidang kesehatan, keuangan, pendidikan, dan lain-lain, sudah mulai menerapkan AI dalam proses analisis datanya.

Di dalam bidang pendidikan saat ini di kenal dengan *Educatioan data mining* (EDM) adalah bidang penelitian antar disiplin ilmu baru yang berhubungan dengan pengembangan metode untuk mengeksplorasi data yang berasal dari konteks pendidikan (Ang et al., 2020). EDM menggunakan pendekatan komputasi untuk menganalisis data di bidang pendidikan. Terjadinya peningkatan dalam penggunaan sistem informasi, yang ditandai dengan banyaknya bermunculan berbagai sistem informasi yang digunakan untuk mengolah data seperti administrasi, pengajaran, nilai, keuangan, dan lain-lain, sehingga hal ini menciptakan gudang

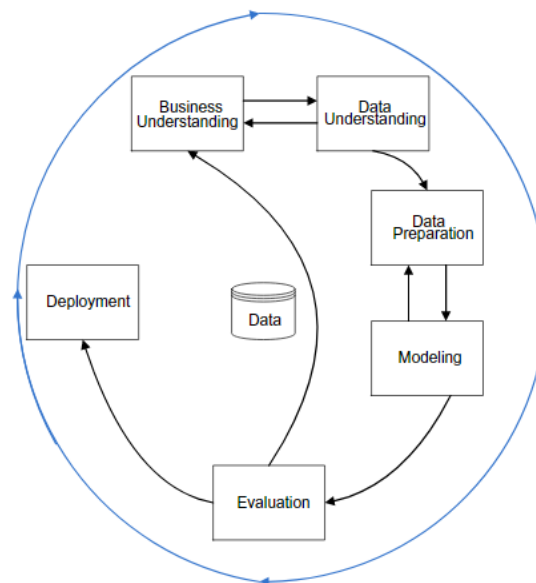
data yang besar yang perlu penanganan khusus agar data-data yang banyak tersebut bisa digunakan untuk menghasilkan suatu yang jauh lebih bermanfaat daripada hanya menjadi tumpukan data saja. Gudang data pendidikan menjadi aset pendidikan yang strategis, sangat signifikan dalam memajukan reformasi pendidikan (Romero & Ventura, 2010).

Di universitas metamedia salah satu syarat untuk bisa menyelesaikan studinya adalah lulus ujian labor profesional. Dimana ujian labor profesional ini menjadi syarat untuk bisa melakukan pendaftaran ujian komprehensif, jika mahasiswa belum lulus ujian labor profesional maka mahasiswa tidak diperbolehkan untuk melakukan pendaftaran ujian komprehensif. Pada saat mahasiswa mengikuti ujian labor profesional banyak mahasiswa yang mengalami kegagalan sehingga harus melakukan beberapa kali ujian profesional. Hal ini berakibat pada keterlambatan mahasiswa dalam menyelesaikan studinya. Tentu hal ini tidak bisa dibiarkan begitu saja, sudah beberapa cara yang dilakukan agar mahasiswa bisa lulus ujian profesional dengan lancar. Seperti dengan diberikannya pelatihan-pelatihan sebelum dilakukan ujian profesional, namun tentu itu tidak bisa menjawab pertanyaan kenapa bisa terjadi kegagalan mahasiswa saat ujian profesional.

Di universitas metamedia sudah banyak sistem informasi yang digunakan untuk mengolah data-data mahasiswa. Dengan adanya gudang data tersebut maka penulis mencoba untuk melakukan analisis penyebab terjadinya kegagalan ujian labor profesional menggunakan salah satu algoritma machine learning yaitu Naïve Bayes (NB). NB adalah salah satu dari 10 algoritma teratas karena kesederhaan, efisiensi, dan kemanjurannya (Zhang & Jiang, 2022).

B.Metodologi Penelitian

Metode penelitian yang digunakan adalah Cross Industry Standard Process Data Mining (CRISP-DM) yaitu model proses independen industri untuk data mining, yang terdiri dari enam fase.



Gambar 1. CRISP-DM

Business Understanding (BU)

Situasi bisnis harus dinilai untuk mendapatkan gambaran tentang sumber daya yang tersedia dan dibutuhkan. Pada fase BU memiliki tugas untuk mengembangkan pemahaman menyeluruh dari perspektif bisnis, informasi latar belakang bisnis, kriteria keberhasilan bisnis dan tentang apa yang benar-benar ingin dicapai.

Data understanding (DU)

Pada fase DU hal yang harus dilakukan adalah mengumpulkan data dari berbagai sumber data, mengeksplorasi dan mendeskripsikannya serta memeriksa kualitas datanya.

Data Preparation (DP)

Pada fase DU dilakukan pemilihan data dengan menentukan kriteria inklusi dan eksklusi. Kualitas data yang buruk dapat ditangani dengan pembersihan data dengan beberapa teknik

pembersihan data seperti data *cleaning* yang bertujuan untuk menyeleksi data dan membuat data yang berpotensi mengurangi akurasi machine learning. Data *integration* yang bertujuan untuk menggabungkan data dari berbagai sumber menjadi satu kesatuan yang lebih besar. Data *transformation* yang bertujuan untuk mengubah struktur data, format data, atau nilai data sedemikian rupa sehingga menghasilkan dataset yang sesuai dengan algoritma yang akan digunakan. Data *reduction* bertujuan untuk mengurangi jumlah sampel data yang diambil.

Modeling

Pada fase modeling adalah fase dimana dilakukan pemilihan teknik permodelan, membangun kasus uji dan model. Secara umum, pemilihan teknik tergantung pada masalah bisnis dan datanya.

Evaluation

Pada fase evaluation adalah fase dimana hasilnya diperiksa terhadap tujuan bisnis yang ditetapkan, keefektifan, serta kualitas algoritma yang digunakan.

Deployment

Pada fase deployment adalah fase dimana mendapatkan sebuah model atau laporan akhir atau komponen perangkat lunak.

C.Pembahasan dan Analisa

Adanya kesulitan untuk mengetahui penyebab terjadinya kegagalan mahasiswa dalam pelaksanaan ujian labor profesional universitas metamedia, sehingga berakibat pada keterlambatan dalam pelaksanaan ujian komprehensif.

Untuk melakukan analisis penyebab kegagalan mahasiswa dalam pelaksanaan ujian labor profesional, didapat data dari sistem informasi akademik kampus universitas metamedia, yang terdiri dari data mahasiswa, data jadwal ujian labor, data nilai ujian labor, data matakuliah, data nilai matakuliah.

Penelitian ini menggunakan data yang diambil dari sistem informasi akademik kampus universitas metamedia. Dimana terdapat data jadwal pelaksanaan ujian labor profesional beserta dengan peserta dan nilai ujian labor pada setiap semester. Ujian labor profesional pada setiap semester dilaksanakan sebanyak 6 kali, sehingga bagi mahasiswa yang belum lulus bisa mengikuti ujian kembali. Kesempatan maksimal untuk bisa ikut ujian labor profesional pada setiap semesternya adalah 6 kali, jika mahasiswa masih belum lulus juga berarti mahasiswa harus menunggu semester berikutnya agar dibuka kembali ujian labor profesional. Hal ini tentu harus bisa diantisipasi dengan metode yang baik, agar mahasiswa bisa menyelesaikan studi tepat waktu.

Selain data jadwal pelaksanaan ujian labor, data lain yang diolah adalah data nilai mahasiswa selama mengikuti perkuliahan di universitas metamedia. Ada 6 nilai matakuliah yang diolah yaitu matakuliah analisa dan perancangan sistem 1, analisa dan perancangan sistem 2, perancangan database, pemrograman 1, pemrograman 2, dan pemrograman berorientasi objek. Profil lulusan prodi sistem informasi universitas metamedia adalah sistem analis dan database administrator, sehingga pada saat ujian labor profesional yang menjadi soal ujian adalah terkait dengan analisa sistem dan pembuatan sebuah sistem.

Nilai yang didapat oleh mahasiswa pada setiap matakuliah di konversikan sehingga menjadi sebuah data set. Konversi nilai matakuliah dilakukan dengan melakukan pengelompokan.

Tabel 1. Konversi Nilai Matakuliah

| Nilai Huruf | Nilai Baru |
|--------------------|-------------------|
| A, A- | Sangat Baik |
| B+, B, B- | Baik |
| C+, C, C-, D, E | Rendah |

Selain nilai matakuliah, hasil kelulusan juga di lakukan konversi, disini penulis melakukan pengelompokan berdasarkan jumlah keikutsertaan dalam melaksanakan ujian labor profesional, jika 1 kali langsung lulus berarti keterangannya lancar, tapi jika lebih dari 1 kali berarti keterangannya tidak lancar.

Tabel 2. Konversi Hasil Kelulusan

| Jumlah Keikutsertaan | Keterangan |
|----------------------|--------------|
| 1 | Lancar |
| >1 | Tidak Lancar |

Algoritma machine learning yang digunakan untuk melakukan analisis prediksi kelulusan ujian labor profesional ini adalah algoritma naïve bayes. Idenya dikembangkan berdasarkan probabilitas bayesian (Alizadeh et al., 2021). Probabilitas bayesian adalah teori probabilitas untuk menghitung besarnya peluang terhadap suatu kejadian yang mana peluang terhadap kejadian tersebut dapat terus berubah sesuai fakta-fakta baru (Chen et al., 2020). Secara umum teorema bayes (Wu et al., 2021) dinyatakan sebagai

$$P(c \vee x_i) = \frac{P(x_i|c)P(c)}{P(x_i)}$$

Dalam notasi ini $P(c \vee x_i)$ berarti peluang kejadian c bila x_i terjadi dan $P(x_i|c)$ peluang kejadian x_i bila c terjadi. Nilai $P(x_i|c)$ disebut likelihood function yang mendeskripsikan peluang untuk pengetahuan fakta-fakta baru terhadap kejadian. Nilai suatu probabilitas harus lebih besar sama dengan nol sampai kurang dari atau sama dengan satu ($0 \leq P \leq 1$). Nilai nol berarti suatu kejadian tidak mungkin muncul, sementara nilai satu berarti suatu kejadian pasti terjadi.

Tahapan dalam algoritma naïve bayes:

1. Hitung probabilitas bersyarat/likelihood

$$P(x_i|c) = \prod_{k=1}^n P(x_k \vee C_j)$$

2. Hitung probabilitas prior untuk tiap class

$$P(c) = \frac{q_j}{q}$$

3. Pilih probabilitas posterior

$$P(c \vee x_i) = \frac{P(x_i|c)P(c)}{P(x_i)}$$

Data yang digunakan sebagai data training sebanyak 315 record. Data training dilakukan perhitungan frekuensi nilai pada setiap atribut dan juga probabilitas setiap nilai atribut.

Tabel 3. Frekuensi nilai atribut keterangan

| Keterangan | Jumlah |
|--------------|--------|
| Lancar | 186 |
| Tidak Lancar | 129 |

Tabel 4. Frekuensi nilai atribut apdsi1

| Keterangan | Lancar | Tidak Lancar |
|-------------|--------|--------------|
| Sangat Baik | 99/186 | 62/129 |
| Baik | 71/186 | 56/129 |
| Rendah | 16/186 | 11/129 |

Tabel 5. Frekuensi nilai atribut apdsi2

| Keterangan | Lancar | Tidak Lancar |
|-------------|---------|--------------|
| Sangat Baik | 123/186 | 63/129 |
| Baik | 52/186 | 56/129 |
| Rendah | 11/186 | 10/129 |

Tabel 6. Frekuensi Nilai Atribut Pemrograman Berorientasi Objek

| Keterangan | Lancar | Tidak Lancar |
|-------------|--------|--------------|
| Sangat Baik | 88/186 | 54/129 |
| Baik | 92/186 | 67/129 |
| Rendah | 6/186 | 8/129 |

Tabel 7. Frekuensi Nilai Atribut Pemrograman 1

| Keterangan | Lancar | Tidak Lancar |
|-------------|--------|--------------|
| Sangat Baik | 97/186 | 54/129 |
| Baik | 63/186 | 52/129 |
| Rendah | 26/186 | 23/129 |

Tabel 8. Frekuensi Nilai Atribut Promrograman 2

| Keterangan | Lancar | Tidak Lancar |
|-------------|---------|--------------|
| Sangat Baik | 125/186 | 69/129 |
| Baik | 51/186 | 47/129 |
| Rendah | 10/186 | 13/129 |

Tabel 9. Frekuensi Nilai Atribut pemrograman basis data

| Keterangan | Lancar | Tidak Lancar |
|-------------|--------|--------------|
| Sangat Baik | 99/186 | 60/129 |
| Baik | 68/186 | 61/129 |
| Rendah | 19/186 | 8/129 |

Setelah mendapatkan pola dari hasil perhitungan data training, maka selanjutnya dilakukan uji coba pada data testing. Data yang digunakan sebagai data testing sebanyak 70 record. Setiap data testing dilakukan perhitungan probabilitas bersyaratnya.

$$\text{Likelihood (Ket = lancar)} = P(\text{lancar}) P(\text{Rendah} | \text{lancar}) P(\text{Sangat Baik} | \text{Lancar}) P(\text{Rendah} | \text{Lancar}) P(\text{Sangat Baik} | \text{Lancar}) P(\text{Sangat Baik} | \text{Lancar}) P(\text{Sangat Baik} | \text{Lancar})$$

$$\frac{186}{315} * 99$$

$$\frac{186}{315} * 123$$

$$\frac{186}{315} * 92$$

$$\frac{186}{315} * 97$$

$$\frac{186}{315} * 125$$

$$\frac{186}{315} * 68$$

$$\frac{186}{315}$$

$$= 0,0131$$

$$\text{Likelihood (Ket = tidak lancar)} = P(\text{tidak lancar}) P(\text{Rendah} | \text{tidak lancar}) P(\text{Sangat Baik} | \text{tidak Lancar}) P(\text{Rendah} | \text{tidak Lancar}) P(\text{Sangat Baik} | \text{tidak Lancar}) P(\text{Sangat Baik} | \text{tidak Lancar}) P(\text{Sangat Baik} | \text{tidak Lancar})$$

$$\frac{129}{315} * 62$$

$$\frac{129}{315} * 63$$

$$\frac{129}{315} * 67$$

$$\frac{129}{315} * 54$$

$$\frac{129}{315} * 69$$

$$\frac{129}{315} * 61$$

$$\frac{129}{315}$$

$$= 0,0052$$

Setelah melakukan perhitungan probabilitas bersyaratnya, maka langkah selanjutnya adalah melakukan perhitungan probabilitas posteriornya.

$$P(\text{Keterangan} = \text{Lancar}) = \frac{\text{likelihood}(\text{ket}=\text{lancar})}{\text{likelihood}(\text{ket}=\text{lancar})+\text{likelihood}(\text{ket}=\text{tidaklancar})}$$

$$\frac{0.0131}{0.0131+0.0052}$$

$$= 0.7136$$

$$P(\text{Keterangan} = \text{Tidak Lancar}) = \frac{\text{likelihood}(\text{ket}=\text{tidaklancar})}{\text{likelihood}(\text{ket}=\text{lancar})+\text{likelihood}(\text{ket}=\text{tidaklancar})}$$

$$\frac{0.0052}{0.0131+0.0052}$$

$$= 0.2863$$

$P_{\text{assignmet}}(\text{ket} = \text{tidak lancar}) < P_{\text{assignmet}}(\text{ket} = \text{lancar})$ maka dihasilnya adalah lancar. Untuk menentukan data tersebut lancar atau tidak lancar adalah dari nilai posterior yang paling besar. Pada kasus data ini prediksinya adalah tepat. Maka dilakukan perhitungan yang sama pak semua data testing.

Setelah dilakukan perhitungan pada semua data traning, maka dilakukan evaluasi akurasi machine learning yang digunakan untuk prediksi kelulusan ujian labor profesional. Metode yang digunakan untuk melakukan evaluasi adalah confusion matrix.

Tabel 10. Hasil Pengujian Naïve Bayes

| | True Lancar | True Tidak Lancar | Class Precision |
|--------------------|-------------|-------------------|-----------------|
| Pred. Lancar | 38 | 17 | 69.09 % |
| Pred. Tidak Lancar | 5 | 10 | 66.67 % |
| Class recal | 88.37 % | 37.04 % | |

Dari hasil pengujian sistem yang telah dilakukan diketahui bahwa nilai akurasi yang ada menghasilkan 68.57.

D.Penutup

Dengan dilakukan pengolahan data yang didapat dari sistem informasi akademik kampus. Dimana data yang diolah adalah data nilai pelaksanaan ujian labor, yang dibandingkan dengan data nilai yang didapat selama mengikuti perkuliahan di universitas metamedia. Sehingga didapatkan model machine learning untuk melakukan prediksi kelulusan yang dapat digunakan untuk mendeteksi kegagalan yang akan dialami oleh mahasiswa jika melakukan pendaftaran ujian labor. Sehingga nantinya mahasiswa harus memperbaiki kemampuannya terlebih dahulu sebelum melakukan pendaftaran ujian labor profesional.

Daftar Pustaka

- Alizadeh, S. H., Hediehloo, A., & Harzevili, N. S. (2021). Multi independent latent component extension of naive Bayes classifier. *Knowledge-Based Systems*, 213. <https://doi.org/10.1016/j.knosys.2020.106646>
- Ang, K. L. M., Ge, F. L., & Seng, K. P. (2020). Big Educational Data Analytics: Survey, Architecture and Challenges. In *IEEE Access* (Vol. 8, pp. 116392–116414). Institute of Electrical and Electronics Engineers Inc. <https://doi.org/10.1109/ACCESS.2020.2994561>
- Chen, S., Webb, G. I., Liu, L., & Ma, X. (2020). A novel selective naïve Bayes algorithm ☆. 192, 105361. <https://doi.org/10.1016/j.knosys>
- Romero, C., & Ventura, S. (2010). Educational data mining: A review of the state of the art. In *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews* (Vol. 40, Issue 6, pp. 601–618). <https://doi.org/10.1109/TSMCC.2010.2053532>
- Wu, C., Zhang, Q., Cheng, Y., Gao, M., & Wang, G. (2021). Novel three-way generative classifier with weighted scoring distribution. *Information Sciences*, 579, 732–750. <https://doi.org/10.1016/j.ins.2021.08.025>
- Zhang, H., & Jiang, L. (2022). Fine tuning attribute weighted naive Bayes. *Neurocomputing*, 488, 402–411. <https://doi.org/10.1016/j.neucom.2022.03.020>