



IMPLEMENTASI *DATA MINING* EVALUASI KINERJA PENELITIAN MAHASISWA DENGAN MENGGUNAKAN ALGORITMA *NAIVE BAYES CLASSIFIER*

Tifani Intan Solihati¹, Nur Hidayanti², Raden Kania³
^{1,2,3} Universitas Banten Jaya

Corresponding Author:

Tifani Intan Solihati
Universitas Banten Jaya,
Warung pojok doni, Jl. Ciwaru Raya No.73, Cipare, Kec. Serang, Kota
Serang, Banten 42117
Email: tifaniintansolihati@unbaja.ac.id

Informasi Artikel:

Submited 9 November, 2021
Direvisi 20 Januari, 2022
Diterima 21 Januari, 2022

How to Cite:

Solihati, T. I., Nurhidayanti, N., & Kania, R. (2022). Implementasi Data Mining untuk Evaluasi Kinerja Penelitian Mahasiswa Universitas Banten Jaya Dengan Menggunakan Algoritma Naïve Vayes Classifier. *Jurnal Theorems (The Original Research of Mathematics)*, X(X), XX-XX.

ABSTRAK

Kinerja Penelitian Mahasiswa bersumber dari hasil penelitian mahasiswa berupa skripsi atau TA. Selama ini Universitas Banten Jaya menerapkan pengukuran karya ilmiah mahasiswa melalui pengujian baik berupa seminar atau sidang skripsi saja, namun belum ada penilaian kinerja penelitian mahasiswa yang dilakukan. Penelitian ini bertujuan untuk memberikan rekomendasi penilaian kinerja mahasiswa melalui publikasi. Fokus penelitian ini untuk mengevaluasi kinerja penelitian mahasiswa dari tahun 2019-2020 di Universitas Banten Jaya dengan menggunakan *data mining*. *Data mining* memiliki tujuan untuk memanfaatkan data dalam basis data dengan mengolahnya sehingga menghasilkan informasi baru yang berguna. Pada penelitian ini menggunakan algoritma *Naïve Bayes Classifier* (NBC) untuk membentuk tabel probabilitas sebagai dasar proses klasifikasi kinerja penelitian mahasiswa. Atribut yang digunakan terdapat 5 macam dengan masing-masing memiliki value ya dan tidak. Berdasarkan hasil klasifikasi NBC pada tools weka diperoleh kesimpulan bahwa metode NBC berhasil mengklasifikasikan 50 data dari 64 data yang diuji. Sehingga metode NBC berhasil memprediksi artikel mahasiswa yang sudah terpublish dengan presentase akurasi sebesar 78,125% sementara presentase untuk *incorrectly classified instances* sebesar 21,875% . Hasil dari NBC dapat digunakan untuk mengklasifikasikan kinerja penelitian mahasiswa tahun selanjutnya dan data penelitian ini dapat dijadikan sebagai data training.

Kata kunci: Algoritma, *Naïve Bayes Classifier*, Data Mining, Evaluation, Implementasi.

ABSTRACT

Student Research Performance is sourced from student research results in the form of a thesis or TA. So far, the University of Banten Jaya has applied the measurement of student scientific work through testing either in the form of seminars or thesis hearings, but no assessment of student research performance has been carried out. This study aims to provide recommendations for student performance assessment through publications. The focus of this research is to evaluate the research performance of students from 2019-2020 at the University of Banten Jaya using data mining. Data mining has the goal of utilizing data in a database by processing it so as to produce useful new information. In this study, the Naïve Bayes Classifier (NBC) algorithm is used to form a probability table as the basis for the classification process of student research performance. There are 5 kinds of attributes used, each with a yes and no value. Based on the results of the NBC classification on the weka tools, it was concluded that the NBC method was successful in classifying 50 data from 64 data tested. So the NBC method successfully predicts published student articles with an accuracy percentage of 78.125% while the percentage for incorrectly

classified instances is 21.875%. The results of the NBC can be used to classify student research performance in the following year and this research data can be used as training data..

Keywords: *Algoritma, Naïve Bayes Classifier, Data Mining, Evaluation, Implementation*

PENDAHULUAN

Mahasiswa merupakan aspek penting dalam upaya mencapai keberhasilan untuk evaluasi suatu Perguruan Tinggi. Mulai dari pemantauan mahasiswa yang masuk Perguruan Tinggi, peningkatan kemampuan atau keahlian mahasiswa, prestasi yang dicapai oleh mahasiswa, rasio jumlah mahasiswa yang lulus terhadap seluruh jumlah mahasiswa yang ada, publikasi penelitian mahasiswa, dan kompetensi lulusan seharusnya mendapat perhatian yang sangat serius untuk memperoleh kepercayaan *stakeholder* dalam menilai dan menetapkan Penggunaan lulusannya. Selama ini Universitas Banten Jaya menerapkan pengukuran karya ilmiah mahasiswa melalui pengujian baik berupa seminar atau siding skripsi saja, namun belum ada penilaian kinerja penelitian mahasiswa yang dilakukan. Kinerja penelitian mahasiswa yang terukur akan memudahkan dalam menentukan strategi yang akan dilakukan untuk meningkatkan karya ilmiah mahasiswa dalam upaya meningkatkan kinerja penelitian baik yang dipublikasikan atau pun yang tidak dipublikasi. Berdasarkan uraian diatas, pada penelitian ini akan dibuat suatu aplikasi yang bertujuan untuk memantau evaluasi kinerja penelitian seluruh mahasiswa Universitas Banten Jaya. Pada penelitian ini digunakan *data mining* dengan menggunakan algoritma *Naïve Bayes Classifier* untuk mengklasifikasi kinerja penelitian mahasiswa yang terpublikasi dan tidak terpublikasi. Pada penelitian ini pula dilakukan penyebaran angket berupa google form kepada mahasiswa semester 5 yang sedang melakukan PKN pada Fakultas Ilmu Komputer dan KP pada Fakultas Teknik, semester 6 yang sedang melakukan PPLK pada Fakultas Keguruan dan Ilmu Pendidikan, lalu mahasiswa semester 8 yang sedang melakukan skripsi dan TA. Populasi pada penelitian ini adalah seluruh mahasiswa Universitas Banten Jaya, sedangkan sampel dari penelitian ini adalah mahasiswa yang mengontrak mata kuliah PKN, KP, Skripsi maupun TA. Dalam peningkatan mutu pendidikan, data mining yang merupakan proses semi otomatis untuk mendapatkan pengetahuan dari sekumpulan data dapat digunakan mengevaluasi kinerja penelitian mahasiswa.

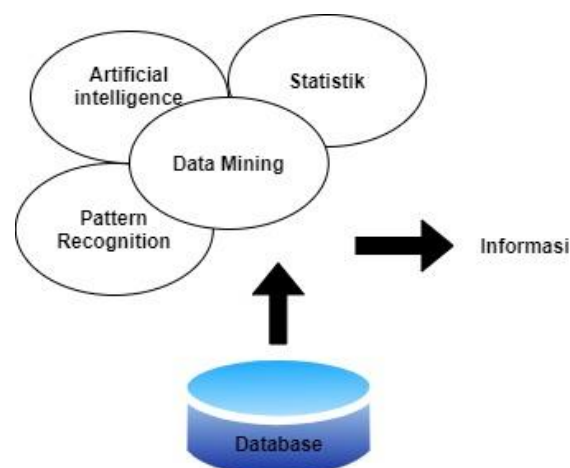
Data mining adalah gabungan dari data statistik, kecerdasan buatan, dan riset basis data yang masih berkembang (Florin Gorunescu, 2011). Data mining juga bisa diartikan sebagai proses yang memperkerjakan satu atau lebih teknik pembelajaran komputer (*mechine learning*) untuk menganalisis dan mengekstaksi pengetahuan (*knowledge*) secara otomatis (Syukri Mustafa et al., 2017). Selain itu, data mining juga bisa diartikan sebagai proses pencarian pola atau informasi menarik dalam data terpilih dengan menggunakan teknik dan metode tertentu (Asroni et al., 2018). Sehingga dapat disimpulkan bahwa data mining merupakan proses pencarian pola data yang tidak diketahui atau tidak diperkirakan sebelumnya. Konsep data mining dapat diterapkan dalam berbagai bidang mislanya bidang pemasaran,

pendidikan, kesehatan, pasar saham, customer relationship management (CRM), teknik, dan lain sebagainya

Beberapa penelitian banyak yang menggunakan data mining untuk menggali beberapa informasi dari sebuah *database* mahasiswa, seperti untuk analisis hubungan antara jalur masuk mahasiswa baru, prestasi dan kelulusan dengan menggunakan algoritma K-Means (Erdoğan & Timor, 2005). Ada pula yang menggunakan data mining untuk memonitoring dan evaluasi kinerja akademik mahasiswa (Ogor, 2007). Lain halnya dengan penelitian yang menghasilkan penerapan data mining dengan teknik klasifikasi yang menghasilkan sebuah pohon kepuasan mengenai kinerja studi mahasiswa selama 5 tahun (Mandias, 2015). Tetapi pada penelitian (Murtopo, 2016) menyebutkan bahwa data proses masuk, asal sekolah, kota asal dan program studi menjadi pertimbangan pada perhitungan tingkat kelulusan. Dan pada penelitian (Mayadewi & Rosely, 2015) menghasilkan data kinerja mahasiswa pada mata kuliah pendukung proyek akhir mereka menggunakan algoritma ID3, CHAID dan Naïve Bayes, bahwa ketiga algoritma tersebut tidak dapat diterapkan untuk nilai proyek.

Aplikasi *data mining* yang kami gunakan dalam penelitian ini adalah *Waikato Environment for Knowledge Analysis* (WEKA). Aplikasi WEKA ini merupakan *open source data mining tool* yang aman digunakan karena mudah digunakan. Berdasarkan penjelasan diatas, kami melakukan implementasi *data mining* untuk evaluasi kinerja penelitian mahasiswa dengan menggunakan algoritma *Naive Bayes Classifier* pada WEKA di Universitas Banten Jaya.

Data mining memiliki tujuan untuk memanfaatkan data dalam basis data dengan mengolahnya sehingga menghasilkan informasi baru yang berguna. Data mining dapat diilustrasikan dengan gambar sebagai berikut:



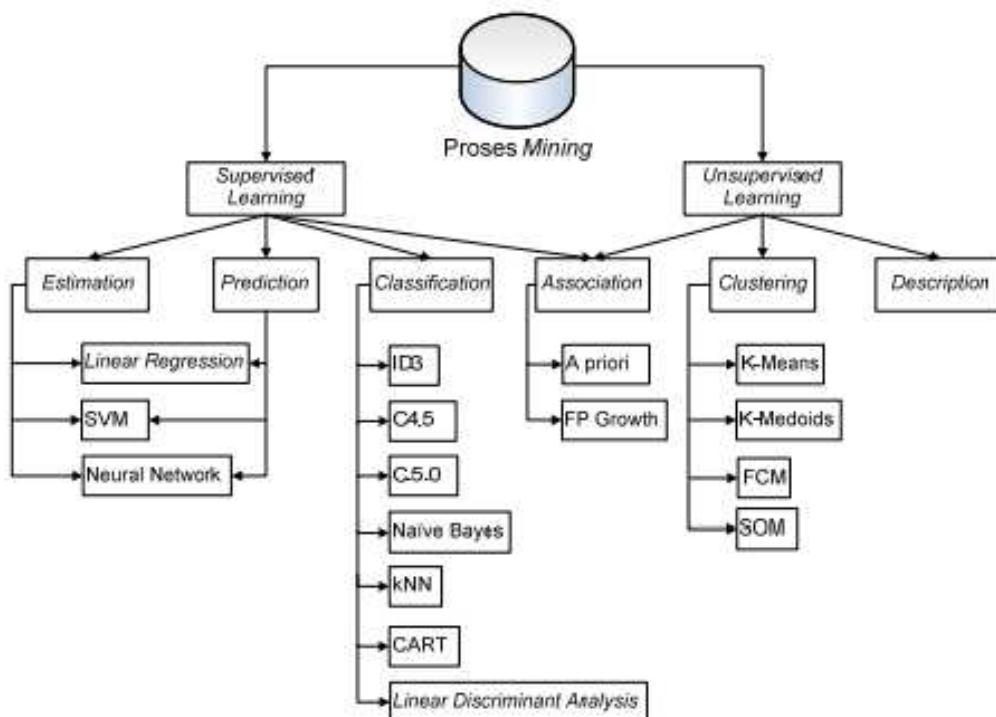
Gambar 1. Akar Ilmu Data Mining (Prasetyo, 2014)

1. Statistik, merupakan akar yang paling tua. Tanda adanya data statistik maka data mining mungkin tidak ada. Data yang diolah dapat diringkas dengan apa yang dikenal sebagai *exploratory data*

analysis (EDA). EDA berguna untuk mengidentifikasi hubungan sistematis antar variabel ketika tidak ada informasi alami yang dibawanya.

2. Kecerdasan Buatan atau *artificial intelligence* (AI). Pada bidang ini, berbeda dengan bidang statistik. Teorinya dibangun pada bidang heuristic sehingga AI berkontribusi terhadap teknik pengolahan informasi berdasarkan pada model penalaran manusia.
3. Pengenalan Pola. Pada tahap ini, sebenarnya data mining juga menjadi turunan bidang dengan pengenalan pola, tetapi hanya mengolah data dari basis data nya saja. Data yang diambil bukan dalam bentuk relasi, melainkan dalam bentuk data normal.
4. Sistem Basis Data. Pada bidang ini, data mining yang menyediakan informasi berupa data yang akan dicari lebih dalam lagi dengan menggunakan metode-metode yang disebutkan sebelumnya.

Data mining memiliki berbagai metode atau fungsi yang dapat digunakan untuk menggali dan menemukan pengetahuan. Terdapat enam kelompok fungsional data mining (Syukri Mustafa et al., 2017), yaitu:



Gambar 2. Beberapa Metode *Data Mining* (Syukri Mustafa et al., 2017)

1. Deskripsi (*description*), memberikan gambaran secara ringkas terhadap sejumlah data yang berskala besar dan memiliki banyak jenis. Termasuk didalamnya metode *Decision Tree*, *Exploratory Data Analysis* dan *Neural Network*.
2. Estimasi (*estimation*), menerka suatu nilai yang belum diketahui, Contohnya menerka penghasilan seseorang ketika beberapa informasi mengenai orang tersebut sudah diketahui. Metode yang dapat

digunakan adalah *Point Estimation*, *Confidance Internal Estimatiois*, *Simple Linear Regression*, *Correlation*, dan *Multiple Regression*.

3. Prediksi (*prediction*), memperkirakan suatu nilai dimasa mendatang, Contohnya memprediksi stok barang tiga tahun kedepan. Yang termasuk fungsi ini antara lain metode *Neural Network*, *Decision Tree*, dan *k-Nearest Neighbor*.
4. Klasifikasi (*classification*), merupakan proses dalam menemukan suatu model atau fungsi yang dapat membedakan konsep atau kelas data, dengan tujuan untuk dapat memperkirakan kelas dari suatu objek yang labelnya tidak diketahui. Yang termasuk dalam fungsi ini antara lain *Neural Network*, *Decision Tree*, *K-nearest Neighbor*, dan *Naïve Bayes*.
5. Pengelompokkan (*clustering*), pengelompokkan untuk mengidentifikasi data dengan karakteristik tertentu. Yang termasuk dalam fungsi ini diantaranya model *Hierarchical Clustering*, metode *k-Means*, dan *Self Organizing Map (SOM)*.
6. Assosiasi (*Association*), biasa disebut juga analisis keranjang pasar dimana fungsi ini digunakan untuk mengidentifikasi item-item produk yang kemungkinan dibeli konsumen bersamaan dengan produk lain. Yang termasuk dalam metode atau algoritma dalam fungsi ini antara lain *Apriori*, *Generalized Sequential Pattern (GSP)*, *FP-Growth* dan *GRI Algorithm*.

Teorema keputusan Bayes adalah pendekatan statistik yang fundamental alam pengenalan pola (*pattern recognition*) (Ridwan et al., 2013). *Naive Bayes Classifier* juga merupakan teknik klasifikasi probabilistic berdasarkan Teori Bayes dengan asumsi independensi yang kuat dalam proses pengklasifikasian (Prasetyowati & Ramadhani, 2018). Ciri utama dari *Naive Bayes Classifier* ini adalah asumsi yang sangat kuat akan independensi dari masing-masing kondisi atau kejadian.

Formula perhitungan *Naive Bayes Classifier* berdasarkan probabilitas ditunjukkan sebagai Berikut:

$$p(A|B) \cdot p(B) = p((B|A)) \cdot p(A)$$
$$p(A|B) = \frac{p(A_i) \cdot p(B|A_i)}{\sum_j^c = 1 p(A_j) \cdot p(B|A_j)}$$

Dengan mengubah nilai A_i dan A_j kedalam vektor "x" maka didapatkan bentuk formula sebagai berikut:

$$p(x|i) = \frac{p(i|x) \cdot p(x)}{\sum_j^c = 1 p(j) \cdot p(x|j)}$$

Adapun perhitungan *Naive Bayes Classifier* untuk data kontinu menggunakan *Distribusi Gauss* sebagai berikut :

$$g(x, \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Keterangan :

$p(x|i)$ = Probabilitas hipotesis x jika diberikan fakta atau *record* i (*Posterior probability*)

$p(i|x)$ = mencari nilai parameter yang memberi kemungkinan yang paling besar (*Likelihood*)

$p(x)$ = *Prior Probability* dari i (*Prior Probability*)

$p(i)$ = Jumlah *probability tuple* yang muncul

g = *Distribusi Gauss*

μ = Rata – rata

σ = Standar Deviasi

Jika $p(i|x)$ dapat diketahui berdasarkan perhitungan diatas, maka kelas/label dari data sampel x adalah kelas (label) yang memiliki $p(i|x) * p(i)$ maksimum.

METODE PENELITIAN

1. Data Mining

Data mining atau penambangan informasi merupakan fitur lunak yang digunakan untuk menciptakan pola tersembunyi, tren, ataupun aturan-aturan yang ada dalam basis yang berdimensi besar serta membuat aturan-aturan yang digunakan buat memperkirakan sikap di masa mendatang (Nikmatun & Waspada, 2019). Data Mining dapat diartikan juga sebagai proses dalam eksplorasi dan analisis data yang dapat dilakukan dengan banyak metode yang masing-masing memiliki kegunaan berbeda. Atau dapat juga diartikan sebagai proses mengekstraksi informasi dari set data besar melalui penggunaan algoritma dan teknik yang diambil dari bidang statistik dan Manajemen Sistem Database (Sukarna Royan, Ansori Yulian, 2021). Klasifikasi dalam informasi mining ialah tata cara pendidikan informasi buat memprediksi nilai dari sekelompok atribut. Algoritma klasifikasi hendak menciptakan sekumpulan ketentuan yang disebut rule yang hendak digunakan selaku penanda buat bisa memprediksi kelas dari informasi yang mau diprediksi (Hozairi et al., 2021).

2. Naïve Bayes

Algoritma Naive Bayes (NB) merupakan metode yang sederhana dalam klasifikasi berdasarkan teori probabilitas yang dikemukakan pertama kali oleh ilmuwan inggris bernama Thomas Bayes. Disebut naif karena menyederhanakan masalah yang bergantung pada dua asumsi penting (Edin Osmanbegović, 2012). Keuntungan dari klasifikasi Naïve Bayes adalah bahwa algoritma ini tidak membutuhkan data pelatihan dalam jumlah yang besar dalam proses klasifikasi (Bhardwaj, Brijesh Kumar, 2011).

Data set yang digunakan pada NBC meliputi sample mahasiswa yang sedang ataupun telah mengontrak mata kuliah PKN/KP/KKM, dan Skripsi yang ada di setiap jurusan Tahun Ajaran 2019-2020 Genap pada Universitas Banten jaya. Mahasiswa tersebut mengisi kuisisioner yang dibagikan melalui *link google form* sebanyak 66 *record*. Data set tersebut terdiri dari 12 atribut diantaranya *timestamp*, *email address*, nama lengkap, NIM, prodi, semester, saran, dan sebagainya.

Tahapan selanjutnya pada *data preprocessing* kami melakukan *data cleansing* untuk membuang data yang mempunyai informasi tidak lengkap. new data set merupakan bagian dari data set sebelumnya namun hanya memiliki 5 atribut diantaranya Apakah pihak kampus (Akademik/Fakultas) mewajibkan tiap mahasiswa untuk mempublikasikan PKN/KP/TA/Skripsi? (KW), Apakah dosen pembimbing

mewajibkan mahasiswanya untuk mempublikasikan PKN/KP/TA/Skripsi yang Anda buat? (**PW**), Apakah mahasiswa melakukan pelaporan ke fakultas bahwa PKN/KP/TA/Skripsi telah terpublikasikan? (**LFP**), Apakah Anda melakukan pelaporan ke fakultas bahwa PKN/KP/TA/Skripsi tidak dipublikasikan? (**LFT**), dan Apakah Laporan Akhir PKN/KP/TA/Skripsi yang Anda buat telah terpublikasikan dalam bentuk jurnal/Prosiding? (**PUBLIKASI**). Sehingga banyak data yang kami gunakan 64 record. Kemudian file tersebut dirubah dalam bentuk .arff yang disebut *new data set* yang diolah menggunakan tool WEKA.

File tersebut masuk ke dalam proses *10-fold cross validation*. Himpunan dibagi ke dalam $-k$ himpunan bagian secara acak. Pada tahap ini perulangan dilakukan sebanyak $-k$ kali dan disetiap perulangan disisakan satu subset untuk pengujian dan subset-subset lainnya untuk pelatihan (Sulistyo, Kustiyo, dan Bueno, 2008). Kami menggunakan metode *10-fold cross validation*. Cara ini bekerja data dibagi menjadi subset 10 bagian (9 subsets sebagai training dan 1 subset sebagai testing). Data tersebut di uji kedalam NBC sehingga menghasilkan model yang paling optimal.

Evaluasi dari NBC menggunakan *confusion matrix* yang terdiri dari nilai accuracy, precision, recall, dan F-measure diperoleh dengan rumus (Han & Kamber, 2006) :

$$\text{Precision} = \frac{TP}{TP + FP}$$
$$\text{Recall} = \frac{TP}{TP + FN}$$
$$F - \text{Measure} = \frac{2 \times \text{precision} \times \text{recall}}{TP + FN}$$
$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN}$$

Dimana :

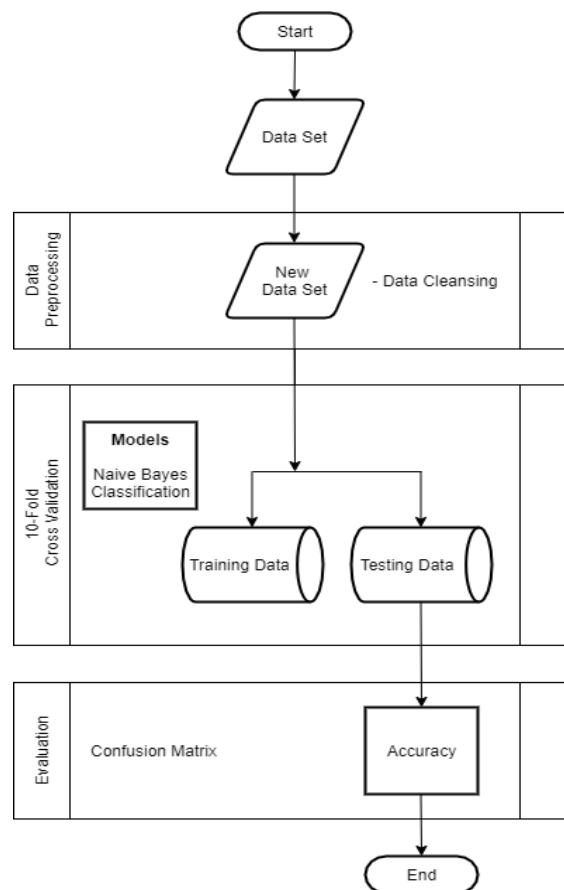
TP atau *True Positives* adalah jumlah data yang dilabeli dengan benar oleh NBC.

TN atau *True Negatives* adalah jumlah data negatif yang dilabeli dengan benar oleh NBC.

FP atau *False Positives* adalah jumlah data negatif yang salah dilabeli oleh NBC.

FN atau *False Negatives* adalah jumlah tuple positif yang salah dilabeli oleh NBC.

Selengkapnya dapat di lihat pada diagram alir dibawah ini:



Gambar 3. Diagram Alir Data Mining untuk Algoritma *Naïve Bayes Classifier*

HASIL DAN PEMBAHASAN

Implementasi WEKA

Weka merupakan sebuah perangkat lunak yang memiliki banyak algoritma machine learning untuk keperluan data mining. Weka juga memiliki banyak tools untuk pengolahan data, mulai dari *pre-processing*, *classification*, *regression*, *clustering*, *association rules*, dan *visualization*. Weka juga merupakan perangkat lunak *open source* berbasis Java dan kita dapat menggunakannya secara langsung atau melalui program Java kita. Weka juga bisa diimplementasikan ke program *python* (Firdaus & Mukhlis, 2020). Weka menelusuri karakteristik atribut dari inputan file *.arff* untuk menelusuri karakter atribut pada proses awal. Atribut yang digunakan KW, PW, LFp, Lft, dan Publikasi masing-masingnya memiliki nilai value ya, tidak, seperti pada gambar di bawah ini:


```

PublikasiDataset - Notepad
File Edit Format View Help
@relation PublikasiDataset

@attribute KW {Ya, Tidak}
@attribute PW {Ya, Tidak}
@attribute LFp {Ya, Tidak}
@attribute Lft {Ya, Tidak}
@attribute Publikasi {ya, Tidak}

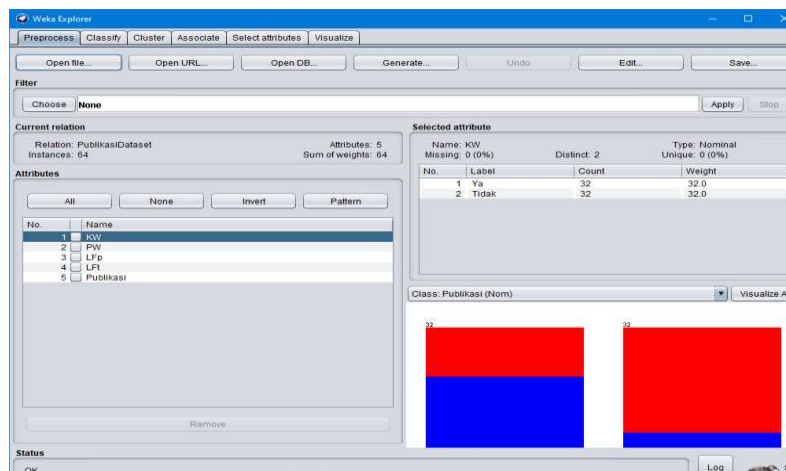
@data

Tidak Tidak Ya Tidak ya
Tidak Tidak Tidak Tidak Tidak
Ya Ya Ya Ya Tidak
Ya Tidak Ya Tidak Tidak
Tidak Tidak Tidak Tidak Tidak
Ya Ya Ya Tidak ya
Tidak Tidak Tidak Tidak Tidak
Tidak Tidak Tidak Tidak Tidak
Tidak Tidak Tidak Tidak Tidak
Tidak Tidak Tidak Tidak Tidak
Tidak Tidak Tidak Tidak Tidak

```

Gambar 4. New Data Set

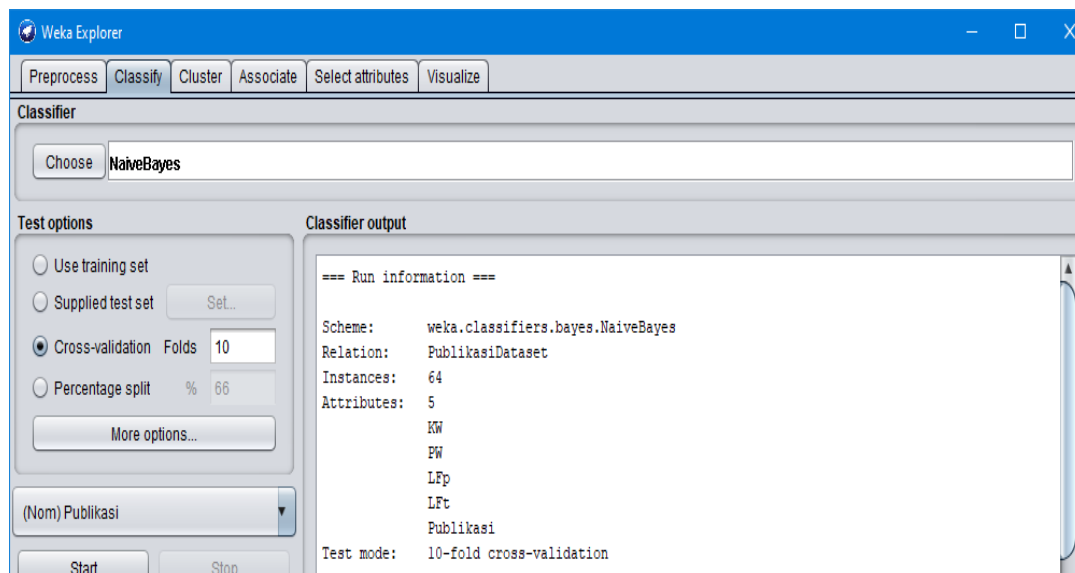
New data set tersebut di import ke tools weka masuk data preprocessing, yang dimana data sudah dibersihkan sebelumnya sehingga muncul seperti gambar di bawah ini. Lima atribut digunakan pada proses pengklasifikasian.



Gambar 5. Data Preprocessing

Gambar 5 merupakan hasil dari atribut KW, PW, LFp, Lft, dan publikasi diimplementasikan pada Weka. Terlihat pada gambar tersebut histogram weka terdapat warna biru dan merah. Histogram sebelah kiri merupakan atribut KW yang sedang ditampilkan dengan masing-masing *value* ya dan tidak terapat 32 *record*. Histogram sebelah kanan adalah target record yang terdapat pada atribut PUBLIKASI dengan *value* ya dan tidak yang akan menjadi kelas target pada penelitian ini.

1. *Data Preparation* atau bisa disebut juga dengan data *preprocessing* adalah suatu proses/langkah yang dilakukan untuk membuat data mentah menjadi data yang berkualitas (input yang baik untuk data mining tools).
2. Pemodelan adalah tahapan dalam membuat model dari suatu sistem nyata.
3. WEKA merupakan *software* terintegrasi yang berisi implementasi dari metode-metode data mining.
4. 10 Fold-Cross validation untuk membantu menghasilkan tingkat keakurasian berdasarkan dataset kualitatif prediksi kebangkrutan yang telah dilakukan klasifikasi.



Gambar 6. Deklarasi Pengujian Klasifikasi NBC pada Weka

Berdasarkan gambar 6. New data set selanjutnya akan melalui tahapan klasifikasi NBC dengan 10 *fold cross validation* dan merupakan implementasi atribut pada weka

```

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      50           78.125 %
Incorrectly Classified Instances    14           21.875 %
Kappa statistic                    0.5424
Mean absolute error                 0.259
Root mean squared error             0.4482
Relative absolute error             56.026 %
Root relative squared error         93.2909 %
Total Number of Instances          64

=== Detailed Accuracy By Class ===

```

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0,783	0,220	0,667	0,783	0,720	0,547	0,773	0,602	ya
	0,780	0,217	0,865	0,780	0,821	0,547	0,773	0,818	Tidak
Weighted Avg.	0,781	0,218	0,794	0,781	0,784	0,547	0,773	0,740	

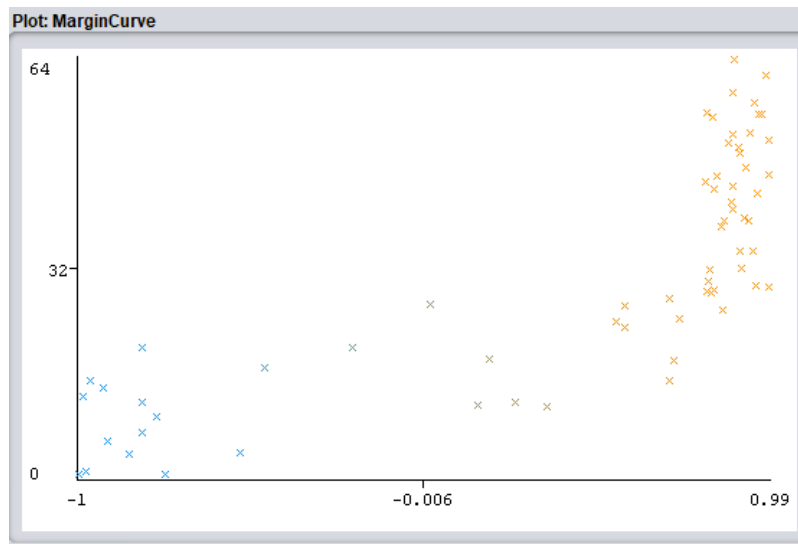
```

=== Confusion Matrix ===
 a b  <-- classified as
18 5 | a = ya
 9 32 | b = Tidak

```

Gambar 7. Hasil Confusion Matrik

Berdasarkan hasil gambar 7. *confusion matrik* dapat dilihat untuk *correctly classified instances/accuracy* sebesar 78,125% sementara presentase untuk *incorrectly classified instances* sebesar 21,875%. Data artikel mahasiswa dari 64, ada 50 data artikel berhasil diklasifikasikan dengan benar sebanyak 50 data dan sebanyak 14 data artikel mahasiswa yang tidak berhasil di klasifikasikan dengan benar. Untuk lebih jelas dapat melihat grafik hasil klasifikasi NBC pada grafik dibawah ini.



Gambar 8. Grafik Hasil Klasifikasi NBC

Grafik di atas merupakan hasil dari pengklasifikasian NBC pada pengujian ini tersebar menjadi 3 area, Grafik yang cenderung ke kiri/yang berwarna biru merupakan data yang terpublikasi dan benar diklasifikasikan. Sedangkan yang sebelah kanan kelompok yang berwarna jingga adalah data yang tidak terpublikasi dan benar dalam pengklasifikasian. Serta data yang tersebar di tengah adalah data yang tidak berhasil di klasifikasikan oleh NBC baik data yang terpublikasi dan data yang tidak terpublikasi.

KESIMPULAN

Berdasarkan hasil klasifikasi NBC pada tools weka diperoleh kesimpulan bahwa:

1. Metode NBC berhasil mengklasifikasikan 50 data dari 64 data yang diuji. Sehingga metode NBC berhasil memprediksi artikel mahasiswa yang sudah terpublikasi dengan presentase accuracy sebesar 78,125%.
2. Hasil dari NBC dapat digunakan untuk mengklasifikasikan kinerja penelitian mahasiswa tahun selanjutnya dan data penelitian ini dapat dijadikan sebagai data training.

SARAN

Berikut saran untuk pengembangan yang dapat dilakukan pada penelitian klasifikasi kinerja penelitian mahasiswa :

1. Atribut yang digunakan dapat ditambah seperti tugas yang terpublikasi.

2. Terdapat beberapa algoritma yang dapat digunakan untuk pengklasifikasian, kami menyarankan menggunakan algoritma lain atau menggabungkan dua algoritma klasifikasi agar mendapat hasil yang lebih baik.

UCAPAN TERIMA KASIH

Terima kasih kepada Tuhan Yang Maha Kuasa, berkat anugerah-Nya lah peneliti dapat melaksanakan penelitian ini, tak lupa ucapan terima kasih penulis kepada penyandang dana penelitian Ditjen Penguatan Riset dan Pengembangan Kementerian Pendidikan, Kebudayaan, Riset dan Teknologi pada skema Penelitian Dosen Pemula (PDP) Tahun Anggaran 2021. Terima kasih pula kepada pihak yang terkait, yakni Lembaga Peneliti dan Pengabdian Masyarakat di Universitas Banten Jaya yang telah menjadi mediator sekaligus lembaga yang mem^{support} penuh atas penelitian para dosen di Universitas Banten Jaya

DAFTAR PUSTAKA

- Aroni, A., Fitri, H., & Prasetyo, E. (2018). Penerapan Metode Clustering dengan Algoritma K-Means pada Pengelompokan Data Calon Mahasiswa Baru di Universitas Muhammadiyah Yogyakarta (Studi Kasus: Fakultas Kedokteran dan Ilmu Kesehatan, dan Fakultas Ilmu Sosial dan Ilmu Politik). *Semesta Teknika*, 21(1), 60–64. <https://doi.org/10.18196/st.211211>
- Bhardwaj, Brijesh Kumar, S. P. (2011). Data Mining: A prediction for performance improvement using classification. *International Journal of Computer Science and Information Security*, 9(4).
- Edin Osmanbegović, M. S. (2012). DATA MINING APPROACH FOR PREDICTING STUDENT PERFORMANCE. *Journal of Economics and Business*, X(1).
- Erdoğan, Ş. Z., & Timor, M. (2005). a Data Mining Application in a Student Database. *Journal of Aeronautics & Space Technologies / Havacilik ve Uzay Teknolojileri Dergisi*, 2(2), 53–57. <http://search.ebscohost.com/login.aspx?direct=true&AuthType=cookie,ip,uid&db=afh&AN=29406812&site=ehost-live>
- Firdaus, F., & Mukhlis, A. (2020). Implementasi Algoritma Naive Bayes Pada Data Set Kualitatif Prediksi Kebangkrutan. *JURIKOM (Jurnal Riset Komputer)*, 7(1), 15. <https://doi.org/10.30865/jurikom.v7i1.1757>
- Florin Gorunescu. (2011). *Intelligent Systems Reference Library* (12th ed.). Springer. <https://doi.org/10.1007/978-3-642-19721-5>
- Hozairi, H., Anwari, A., & Alim, S. (2021). Implementasi Orange Data Mining Untuk Klasifikasi Kelulusan Mahasiswa Dengan Model K-Nearest Neighbor, Decision Tree Serta Naive Bayes. *Network Engineering Research Operation*, 6(2), 133. <https://doi.org/10.21107/nero.v6i2.237>
- Mandias, G. F. (2015). *Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Di*



Universitas Klabat Dengan Metode Klasifikasi. 9–10.

- Mayadewi, P., & Rosely, E. (2015). Prediksi Nilai Proyek Akhir Mahasiswa Menggunakan Algoritma Klasifikasi Data Mining. *Seminar Nasional Sistem Informasi Indonesia, November*, 329–334.
- Murtopo, A. A. (2016). Prediksi Kelulusan Tepat Waktu Mahasiswa STMIK YMI Tegal Menggunakan Algoritma Naïve Bayes. *CSRID (Computer Science Research and Its Development Journal)*, 7(3), 145. <https://doi.org/10.22303/csrid.7.3.2015.145-154>
- Nikmatun, I. A., & Waspada, I. (2019). Implementasi Data Mining untuk Klasifikasi Masa Studi Mahasiswa Menggunakan Algoritma K-Nearest Neighbor. *Jurnal SIMETRIS*, 10(2), 421–432.
- Ogor, E. N. (2007). Student Academic Performance Monitoring and Evaluation Using Data Mining Techniques Department of Natural Sciences Turks & Caicos Islands Community College Visualization and Articulation. *Fourth Congress of Electronics, Robotics and Automotive Mechanics*, 354–359. <https://doi.org/10.1109/CERMA.2007.78>
- Prasetyowati, E., & Ramadhani, N. (2018). Sistem Evaluasi Dan Klasifikasi Kinerja Akademik Mahasiswa Universitas Madura Menggunakan Naive Bayes Dengan Dirichlet Smoothing. *JUTI: Jurnal Ilmiah Teknologi Informasi*, 16(2), 192. <https://doi.org/10.12962/j24068535.v16i2.a688>
- Ridwan, M., Suyono, H., & Sarosa, M. (2013). *Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier*. 7(1), 59–64.
- Sukarna Royan, Ansori Yulian, S. (2021). Implementasi Data Mining Menggunakan Metode Naive Bayes Dengan Feature Selection Untuk Prediksi Kelulusan Mahasiswa Tepat Waktu. *SAINTEK | Jurnal Sains & Teknologi*, 5(2), 9–22.
- Syukri Mustafa, M., Rizky Ramadhan, M., & Thenata, A. P. (2017). Implementasi Data Mining untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier. *Citec Journal*, 4(2), 151–162.