

IMPLEMENTASI DATA MINING UNTUK MENGLASIFIKASI HASIL KELULUSAN MADRASAH MENGGUNAKAN ALGORITMA NAÏVE BAYES

Teguh Adrian¹, Nana Suarna²

¹ Manajemen Informatika STMIK IKMI Cirebon

² Teknik Informatika STMIK IKMI Cirebon

Jl Perjuangan No 10B Majasem Kesambi Kota Cirebon

adrianteguh090@gmail.com

ABSTRAK

Madrasah merupakan lembaga pendidikan yang sangat penting untuk mempersiapkan generasi bangsa dan agama. Madrasah muncul di Indonesia sejak awal abad ke-20 sebelum Indonesia mengalami kemerdekaan. Perkembangan Pendidikan islam di Indonesia semakin pesat dimulai pada masa awal kemerdekaan Indonesia pada masa ini madrasah sangat jauh berkembang, dengan meningkatnya perkembangan sekolah di Indonesia, indeks yang menjadi syarat kelulusan bagi para siswa juga mengalami keanaikan. Metode yang digunakan dalam Artikel ini yaitu Knowledge Discovery in Database (KDD) dengan tahapan selection, preprocessing, transformation, data mining dan evaluation, algoritma naïve bayes digunakan karena cocok dengan klasifikasi binary yaitu dengan mengklasifikasi dua kategori lulus dan tidak lulus, Tujuan Artikel ini untuk memudahkan para guru untuk mengetahui hasil dari prediksi kelulusan siswa madrasah, kemudian untuk siswa yang nilainya dibawah kriteria kelulusan akan tetap lulus namun bersyarat apakah ada tugas tambahan atau program pengulangan agar mencapai standar kelulusan dan memberikan rekomendasi untuk meningkatkan tingkat kelulusan. Hasil dari Artikel ini memperoleh akurasi sebesar 88,24%, Recall sebesar 92,11% dan precision sebesar 87,50% yang diperoleh dari 132 data latih dan 68 data uji dan mendapatkan ROC sebesar 0.922. karena tingkat akurasi diatas 50% ini berarti kinerja naïve bayes dalam permasalahan ini sudah cukup baik.

Kata kunci : Data mining, klasifikasi, naïve bayes, madrasah, KDD

1. PENDAHULUAN

Madrasah merupakan suatu Lembaga Pendidikan formal dibawah naungan kementerian agama yang Sebagian besar merupakan madrasah swasta dan memiliki daya saing yang masih rendah terutama dalam hal kualitas pembelajaran. Menyadari akan hal itu kementerian agama memberikan suatu kebijakan dalam rangka untuk meningkatkan mutu dan daya saing madrasah salah satu contohnya mencanangkan program madrasah riset pada tahun 2013[1]

Di Indonesia sendiri sekolah ada dua macam yaitu sekolah umum dan sekolah agama, sekolah umum seperti SD, SMP, SMA dan sekolah agama seperti MAN, MA, Madrasah dan lainnya. Dengan meningkatnya perkembangan sekolah di Indonesia indeks yang menjadi syarat kelulusan bagi para siswa juga ikut naik [2].

Madrasah berdiri sejak awal abad ke- 20 sebelum masa kemerdekaan Indonesia, berdirinya madrasah di barengi dengan dibentuknya ormas-ormas islam seperti nadhatul ulama, muhamadiyah dan lainnya.[3]

Penelitian terdahulu yang dilakukan oleh penulis diantaranya oleh Edhy Poerwandono dengan judul Penerapan data mining Untuk penilaian Kinerja Karyawan di Pt Riksa Dinar Djaya Menggunakan Metode Naïve Bayes Classification hasil dari penelitian menggunakan algoritma naïve bayes adalah accuracy sebesar 99.44% untuk pengambilan keputusan dalam menilai kinerja karyawan di PT. Riksa Dinar Djaya [4].

Tujuan Artikel ini adalah untuk mengembangkan sebuah model klasifikasi menggunakan Teknik pembelajaran mesin untuk mengklasifikasi

berdasarkan kelulusan siswa dan meningkatkan minat belajar siswa dan mengetahui hasil dari kelulusan para siswa apakah dinyatakan lulus atau belum lulus

Alasan mengambil judul tersebut karena berdasarkan latar belakang yang di atas maka dilakukan Artikel dengan judul yaitu implementasi data mining untuk mengklasifikasi hasil kelulusan madrasah menggunakan algoritma naïve bayes pada artikel ini algoritma naïve bayes dipilih karena memudahkan peneliti dalam menghitung dan menentukan kemungkinan lulus atau tidaknya siswa.

2. TINJAUAN PUSTAKA

2.1. Madrasah

Kata “Madrasah” merupakan kata serapan dari Bahasa Arab yaitu “darasa” yang memiliki arti tempat belajar bagi para pelajar. Zaman sekarang istilah madrasah mungkin sudah menyatu dengan istilah sekolah ataupun perguruan tinggi khususnya perguruan islam, namun ada yang berpendapat bahwa sistem pengajaran di madrasah sama seperti sekolah pada umumnya namun pengajaran tentang agama islam lebih di khususkan ketimbang sekolah biasa.[5]

2.2. Klasifikasi

Klasifikasi adalah proses untuk menentukan suatu Kumpulan data dengan pola atau fungsi-fungsi yang memisahkan antara kelas yang satu dengan kelas yang lainnya. Tujuan dari klasifikasi adalah untuk memprediksi suatu kelas dari objek yang belum di ketahui kelasnya ada beberapa algoritma dalam

klasifikasi contohnya naïve bayes, KNN, C45 dan lainya [6]

2.3. Data Mining

Data mining digunakan guna meningkatkan pemahaman tentang informasi dalam database, data mining merupakan proses pencarian informasi yang menggunakan Teknik statistika, Matematika dan Marchine learning untuk mengidentifikasi informasi yang bermanfaat dari berbagai database tujuan dari data mining sendiri adalah untuk mendapatkan pola yang mungkin memberikan indikasi yang bermanfaat [7].

Istilah data mining dan KDD sering kali digunakan namun ternyata kedua istilah tersebut memiliki konsep yang berbeda tetapi saling berkaitan karena data mining merupakan bagian dari tahapan dalam proses *knowledge Discovery In Database* [8]

2.4. Naïve Bayes

Metode Naïve bayes di kemukakan oleh seorang ilmuwan berkebangsaan inggris yaitu *Thomas Bayes* dengan memprediksikan peluang di masa depan berdasarkan pengalaman di masa sebelumnya sehingga disebut Teorema Bayes [9]

$$P(H|X) = \frac{P(X|H).P(H)}{P(X)}$$

Antara lain:

X: Data dengan class yang belum diketahui

H: Hipotesis data merupakan suatu class spesifik

$P(H|X)$: probabilitas hipotesis H berdasarkan kondisi X (probabilitas proserior)

$P(H)$: Probabilitas hipotesis H (probabilitas prior)

$P(X|H)$: Probabilitas X berdasarkan kondisi pada hipotesis H

$P(X)$: probabilitas X

Pada algoritma *Naïve Bayes* memanfaatkan probabilitas dan statistic untuk menyelesaikan suatu masalah klasifikasi, kelebihan dari metode *naïve bayes* adalah lebih cepat dalam melakukan perhitungan dan data yang dibutuhkan tidak terlalu banyak dalam melakukan klasifikasi ini [10].

2.5. Confusion Matrix

Confusion matriks adalah sebuah tabel yang berguna untuk mendeskripsikan perfoma suatu model marchine learning [11].

		Ground Truth	
		1	0
Prediksi	1	TP	FP
	0	FN	TN

Gambar 1 Confusion Matrix

Confusion matrix merupakan metode yang banyak dipakai untuk menghitung akurasi pengukurannya memiliki istilah sebagai berikut:

- *True Negative* (TN) prediksi negative, real negative

- *True Positive* (TP) prediksi positive, real positive
- *False Negative* (FN) prediksi negative, real positive
- *False Positive* (FP) prediksi positive, real negative

2.6. Acuracy

Akurasi dihasilkan dari perbandingan antara data yang diklasifikasikan secara benar dengan keseluruhan data akurasi menunjukkan kedekatan hasil klasifikasi dengan nilai sebenarnya [12]

$$Acuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100\%$$

2.7. precision

Precision adalah tingkat kerteepatan yang menunjukkan seberapa dekat perbedaan nilai setiap melakukan pengulangan [13] yang merupakan fraksi positif yang diidentifikasi benar positif.

$$Precision = \frac{TP}{TP+FP} \times 100\%$$

2.8. Recall

Recall atau sensitivitas adalah rasio prediksi benar positif dilakukan dengan seluruh data yang benar positif atau mengukur proporsi positif asli yang diprediksi benar positif [14]

$$Recall = \frac{TP}{TP+FN} \times 100\%$$

3. METODE PENELITIAN

3.1. Teknik Pengumpulan Data

Teknik pengumpulan data yang dilakukan guna mendapatkan data yang digunakan untuk membuat artikel, teknik yang digunakan adalah Wawancara Penelitian ini dilakukan dengan menggunakan wawancara kepada guru di madrasah Nurul Anwar untuk mendapatkan data yang diperlukan.

3.2. Sumber data

Sumber data berasal dari madrasah Nurul Anwar yang berarti merupakan data primer, Metode analisis digunakan untuk membuat artikel adalah analisis deskriptif kuantitatif yang digunakan untuk menguji data berdasarkan perhitungan statistika dan matematika yang digunakan pada penelitian ini.

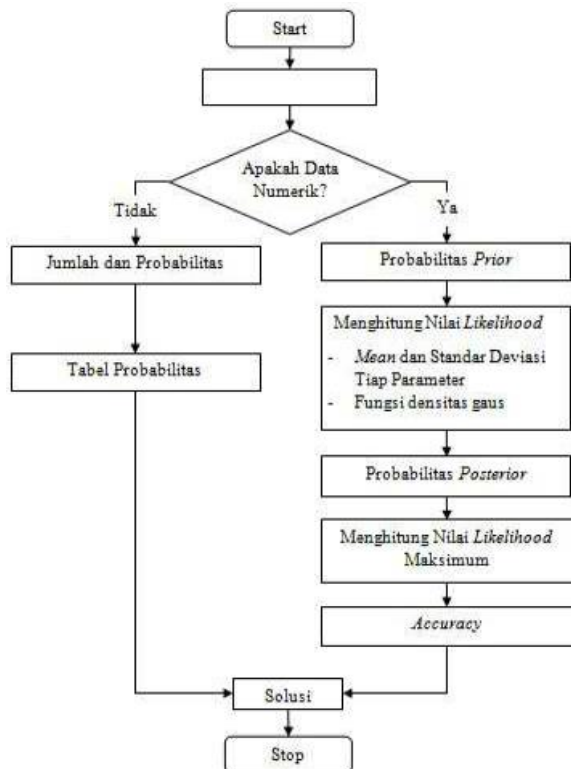
3.3. Metode Naïve bayes

Tahapannya sebagai berikut:

- 1) Mulai Algoritma: Inisialisasi proses Naïve Bayes.
- 2) Persiapkan Data Training
- 3) Hitung Probabilitas Kelas: Hitung probabilitas priori untuk setiap kelas.
- 4) Hitung Probabilitas Atribut untuk Setiap Kelas (Likelihood): Hitung probabilitas likelihood untuk setiap atribut pada setiap kelas.
- 5) Hitung Probabilitas Prior (Priori): Hitung probabilitas priori menggunakan jumlah contoh yang dimasukkan ke dalam perhitungan.
- 6) Hitung Posteriori Probabilitas: Hitung probabilitas posteriori menggunakan rumus

Bayes, yang menggabungkan probabilitas priori dan likelihood.

- 7) Pilih Kelas dengan Probabilitas Tertinggi: Pilih kelas yang memiliki probabilitas posterior tertinggi sebagai kelas yang diprediksi.
- 8) Kelas Terpilih: Kelas yang dipilih menjadi hasil prediksi untuk sampel data tertentu.



Gambar 2 Flowchat Naive bayes

3.4. Metode Analisa Data



Gambar 3 Metode Analisa

Pengembangan *data mining* untuk menganalisis data yaitu menggunakan metode *Knowledge In Database (KDD)* yang tahapannya adalah *data selection*, *pre-processing*, *Transformation*, *Data mining*, *Evaluation*. Adapun tahapannya sebagai berikut:

- a. *Data Selection*
Merupakan sebuah proses menyeleksi data atau menyiapkan data yang akan dijadikan atribut/indicator/sebuah target dalam sebuah data yang besar
- b. *Pre-Processing/Cleaning*
Merupakan sebuah proses untuk memperbaiki kesalahan dalam data, memilah data yang ganda dan data yang dianggap kesalahan cetak.
- c. *Data Transformation*

Merupakan sebuah proses perubahan data. Proses ini dilakukan untuk mempermudah dalam mengolah data dan merupakan data yang sudah siap untuk dilakukan ke tahap *data mining*.

- d. *Data Mining*
Merupakan proses menggali informasi atau pola yang menarik dan bermanfaat dengan menggunakan cara atau metode tertentu.
- e. *Evaluation*
Merupakan sebuah proses menampilkan hasil dan informasi dari proses data mining ke dalam sebuah bentuk yang mudah untuk dimengerti.

4. HASIL DAN PEMBAHASAN

Pada penelitian ini yang dilakukan di Madrasah Nurul Anwar Kecamatan Plumbon Kabupaten Cirebon dengan menggunakan dataset kelulusan pada siswa kelas 5 dari tahun 2019 sampai dengan 2023 dengan jumlah data sebanyak 200 data dengan jumlah data dari tiap tahun masuk sebagai berikut: tahun 2015 sebanyak 35 siswa, 2016 sebanyak 37 siswa, 2017 sebanyak 44 siswa, 2018 sebanyak 51 siswa dan tahun 2019 sebanyak 34 siswa, pada penelitian ini untuk dataset dibagi menjadi dua untuk data training diambil dari setiap kelompok tahun masuk siswa dari tahun 2015 sebanyak 22 siswa, 2016 sebanyak 27 siswa, 2017 sebanyak 29 siswa, tahun 2018 sebanyak 34 siswa dan 2019 sebanyak 21 siswa. Maka total data training sebanyak 132 siswa dan data test sebanyak 68 siswa, dengan menggunakan aplikasi Rapid miner studio menghasilkan hasil sebagai berikut:

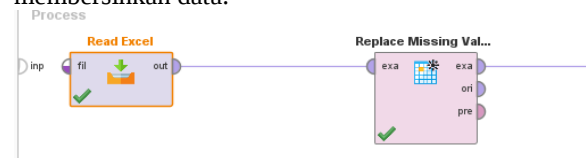
1. Data selection untuk menentukan atribut yang digunakan



Gambar 4 Read Excel

Menggunakan operator read excel untuk membaca dataset kita punya.

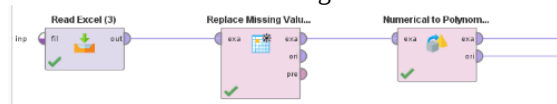
2. Pre-processing/cleaning bertujuan untuk membersihkan data.



Gambar 5 Pre-processing

Pada rapid miner menggunakan operator *replace missing value* dari processing sebanyak 200 data dan semuanya tidak ada data yang *missing* ataupun *noise*

3. Transformation untuk mengubah data.



Gambar 6. Transformation

Operator yang digunakan adalah numerical to polynominal mengubah data numeric menjadi polynominal.

4. Data Mining yang bertujuan untuk menggali informasi dari dataset yang telah kita tentukan Ada operator tambahan yang digunakan yaitu apply model naïve bayes dan performance.



Gambar 7 Data mining

5. Evaluation yang bertujuan untuk menampilkan hasil dari proses penambahan data satu data mining sebagai berikut:

a. Confusion Madrix

		Ground Truth	
		Lulus	Belum Lulus
Prediksi	Lulus	35	5
	Belum Lulus	3	25

Gambar 8 Hasil Confusion

Dihasilkan hasil sebagai berikut

true positive: 35

false positive: 5

true negative: 25

false negative: 3

Acuracy

$$Acuracy = \frac{(35+25)}{(35+5+25+3)} = 0,8824 \times 100\% = 88,24\%$$

Maka didapatkan hasil dari perhitungannya adalah 88,24%

b. Precision

precision@0.95 (positive class: lulus)			
	trial belum lulus	trial lulus	class precision
pred. belum lulus	25	3	88.25%
pred. lulus	5	25	87.50%
class recall	88.25%	92.11%	

Gambar 10 Precision

$$Precision = \frac{35}{(35+5)} = 0,875 \times 100\% = 87,50\%$$

Maka didapatkan hasil dari perhitungannya adalah 87,50%

c. Recall

recall@0.95 (positive class: lulus)			
	trial belum lulus	trial lulus	class precision
pred. belum lulus	25	3	88.25%
pred. lulus	5	25	87.50%
class recall	88.25%	92.11%	

Gambar 11 Recall

$$Recall = \frac{35}{(35+3)} = 0,9211 \times 100\% = 92,11\%$$

Maka didapatkan hasil dari perhitungan adalah 92,11%

d. Roc

Hasil roc sebesar 0.922 itu berarti penggunaan metode naïve bayes sudah cukup bagus dalam mengklasifikasi hasilnya.

e. Confidance



Gambar 12 Roc

hasil Confidancenya sebagai berikut:

Tabel 1 Hasil Confidance

Hasil	Prediction	Confidance (BL)	Confidance (L)	Jenis kelamin	Anal	Tahun Masuk	N Q	N H	N A	N A K	N F	N TI
lulus	lulus	0.01723	0.98276	perempuan	cibinong	2015	9	8	7	7.2	8	7
lulus	lulus	1.54E-04	0.99984	laki-laki	cibinong	2015	8	9	8	8.3	8	7
belum lulus	belum lulus	0.99999	3.24E-06	laki-laki	cibinong	2015	7	7	7	7.3	7	7
belum lulus	belum lulus	0.95067	0.04932	perempuan	cibinong	2015	8	8	7	7.5	8	7
...	...	...	...	...	...	...	...	...	...	...	...	...

hasilnya sebagai berikut untuk tahun 2015 dari 13 data uji terprediksi secara benar semua, untuk tahun 2016 dari 10 data uji terprediksi benar semua, untuk tahun 2017 dari 15 data uji ada 2 data belum lulus terprediksi lulus dan 2 data lulus terprediksi belum lulus, tahun 2018 dari 17 data uji ada 2 data belum lulus ternyata terprediksi lulus dan 1 data lulus ternyata terprediksi belum lulus, tahun 2019 ada 1 data belum lulus ternyata terprediksi lulus.

5. KESIMPULAN DAN SARAN

Jadi dapat disimpulkan bahwa hasil dari implementasi data mining menggunakan metode Analisa Knowledge discovery in database dan algoritma klasifikasi yaitu naïve bayes classifier menghasilkan sebuah informasi baru dan dapat mengklasifikasi hasil nilai kelulusan siswa madrasah dengan membandingkan antara 132 data traning dan 68 data testing dan juga menghasilkan akurasi sebesar 88,24%, recall sebesar 92,11% dan precision sebesar 87,50% dan roc sebesar 0.922.

DAFTAR PUSTAKA

- [1] P. P. Agama, D. Keagamaan, B. Litbang, D. Diklat, and K. Agama, "Website: <http://jurnaledukasikemenag.org> EDUKASI: Jurnal Penelitian Pendidikan Agama dan INOVASI MADRASAH MELALUI PENYELENGGARAAN MADRASAH RISET THE INNOVATION OF MADRASAH THROUGH IT'S RESEARCH ENFORCEMENT Umul Hidayati," vol. 17, no.

- 3, pp. 238–255, 2019, [Online]. Available: <http://jurnaledukasikemenag.org>
- [2] S. Mulyati, S. Maulana Husein, and K. Kunci, “RANCANG BANGUN APLIKASI DATA MINING PREDIKSI KELULUSAN UJIAN NASIONAL MENGGUNAKAN ALGORITMA (KNN) K-NEAREST NEIGHBOR DENGAN METODE EUCLIDEAN DISTANCE PADA SMPN 2 PAGEDANGAN sistem dapat memprediksi dan mengklasifikasikan dengan baik dan cepat,” pp. 65–73, 2020.
- [3] A. T. Hidayatullah, U. Islam, Z. Hasan, G. Probolinggo, J. Mahalli, and A. A. Yaqin, “SEJARAH LEMBAGA PENDIDIKAN ISLAM (MADRASAH) DAN PERKEMBANGANNYA DI INDONESIA,” *Jurnal Mahasiswa*, vol. 4, no. 3.
- [4] E. Poerwandono and J. Perwitosari, “Penerapan Data Mining Untuk Penilaian Kinerja Karyawan Di PT. Riksa Dinar DJaya Menggunakan Metode Naïve Bayes Classification (Edhy Poerwandono 1 , Faizal Joko Perwitosari 2) Penerapan Data Mining Untuk Penilaian Kinerja Karya Di PT Riksa Dinar Djaya Menggunakan Metode Naive Bayes Classification,” *Jurnal Sains dan Teknologi*, vol. 5, no. 1, p. |pp, 2023, doi: 10.55338/saintek.v5i1.1416.
- [5] Y. Chairiyah Madrasah Aliyah Negeri, “SEJARAH PERKEMBANGAN SISTEM PENDIDIKAN MADRASAH SEBAGAI LEMBAGA PENDIDIKAN ISLAM,” 2021.
- [6] A. H. Nasrullah, “IMPLEMENTASI ALGORITMA DECISION TREE UNTUK KLASIFIKASI PRODUK LARIS,” vol. 7, no. 2, 2021, [Online]. Available: <http://ejournal.fikom-unasman.ac.id>
- [7] F. Juliawati, R. Buaton, R. Saragih, and S. Kaputama, “Pengelompokan Data Mining Penerimaan Bantuan Pangan Non Tunai (BPNT) Menggunakan Metode Clustering (Studi Kasus : Kantor Desa Payabakung Hamparan Perak),” 2023.
- [8] H. Nicodemus Turnip and H. Fahmi, “Hendra Turnip, 2 Hasanul Fahmi (Penerapan Data Mining Pada Penjualan Kartu Paket Internet Yang Banyak Diminati Konsumen Dengan Metode K-Means) Penerapan Data Mining Pada Penjualan Kartu Paket Internet Yang Banyak Diminati Konsumen Dengan Metode K-Means,” *JIKOMSI Jurnal Ilmu Komputer dan Sistem Informasi*, vol. 4, no. 2, pp. 36–41, 2021.
- [9] I. Arfanda, W. Ramdhan, R. A. Yusda, and H. Artikel, “Naive Bayes Dalam Menentukan Penerima Bantuan Langsung Tunai,” vol. 1, no. 1, 2021, doi: 10.47709/briliance.vxix.xxxx.
- [10] H. Apriyani, “Perbandingan Metode Naïve Bayes Dan Support Vector Machine Dalam Klasifikasi Penyakit Diabetes Melitus,” 2020. [Online]. Available: <https://journal-computing.org/index.php/journal-ita/index>
- [11] K. Kristiawan and A. Widjaja, “Perbandingan Algoritma Machine Learning dalam Menilai Sebuah Lokasi Toko Ritel,” *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 7, no. 1, Apr. 2021, doi: 10.28932/jutisi.v7i1.3182.
- [12] J. Homepage *et al.*, “MALCOM: Indonesian Journal of Machine Learning and Computer Science Sentiment Analysis of Online Lectures in Indonesia from Twitter Dataset Using InSet Lexicon Analisis Sentimen terhadap Perkuliahan Daring di Indonesia dari Twitter Dataset Menggunakan InSet Lexicon,” vol. 1, pp. 24–33, 2021.
- [13] R. Indra Borman, I. Ahmad, and Y. Rahmanto, “Klasifikasi Citra Tanaman Perdu Liar Berkhasiat Obat Menggunakan Jaringan Syaraf Tiruan Radial Basis Function,” *Bulletin of Informatics and Data Science*, vol. 1, no. 1, 2022, [Online]. Available: <https://ejurnal.pdsi.or.id/index.php/bids/index>
- [14] U. M. Kudus, J. Ganesha, and P. Kudus, “Fida Maisa Hana.”