



IMPLEMENTATION OF DATA MINING TO CLASSIFY MADRASAH GRADUATION RESULTS USING THE NAIVE BAYES ALGORITHM

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

Teguh Adrian¹, Nana Suarna²

^{1,2} Manajemen Informatika Fakultas Teknik Informatika

Sekolah Tinggi Manajemen Informatika dan Komputer IKMI Cirebon

E-mail: adrianteguh090@gmail.com¹, nana.ikmi@gmail.com²

ARTICLE INFO

Correspondent:

Teguh Adrian

adrianteguh090@gmail.com

Key words:

implementation, data mining, classification, madrasah graduation results, naive bayes algorithm

Website:

<https://idm.or.id/JSCR/index.php/JSCR>

Page: 1142 - 1160.

ABSTRACT

Madrasah is a crucial educational institution in preparing the nation's future generations and promoting religious education. Madrasah has been present in Indonesia since the early 20th century, predating the country's independence. The rapid development of Islamic education in Indonesia began during the early years of independence, marked by special attention from the government. Despite improvements, the index of student graduation remains a challenge due to scattered and time-consuming data collection. This Final Project Report utilizes the Knowledge Discovery in Database (KDD) method with the Naïve Bayes Classification algorithm to predict the graduation rates of madrasah students. The goal is to facilitate teachers in understanding the predicted results, provide opportunities for students with grades below the graduation criteria, and offer recommendations to madrasahs to enhance their graduation rates. The results show an accuracy of 88.24%, recall of 92.11%, and precision of 87.50%, indicating the effective performance of the Naïve Bayes algorithm in predicting student graduation.

Copyright © 2023 JSCR. All rights reserved.

INFO ARTIKEL	ABSTRAK
Koresponden Teguh Adrian <i>adrianteguh090@gmail.com</i> Kata kunci: implementasi, data mining, klasifikasi, hasil kelulusan madrasah, algoritma naive bayes Website: <i>https://idm.or.id/JSCR/index.php/JSCR</i> Hal: 1142 - 1160	Madrasah, sebagai lembaga pendidikan penting dalam mempersiapkan generasi bangsa dan agama, telah hadir di Indonesia sejak awal abad ke-20 sebelum kemerdekaan. Perkembangan pendidikan Islam semakin pesat sejak awal kemerdekaan, ditandai dengan perhatian khusus dari pemerintah. Meski mengalami peningkatan, indeks kelulusan siswa masih menjadi tantangan karena data tersebar dan memakan waktu lama dalam pengumpulannya. Laporan Tugas Akhir ini menggunakan metode Knowledge Discovery in Database (KDD) dengan algoritma Naïve Bayes Classification untuk memprediksi tingkat kelulusan siswa madrasah. Tujuannya adalah memudahkan para guru dalam memahami hasil prediksi kelulusan siswa, memberikan peluang bagi siswa dengan nilai di bawah kriteria kelulusan, dan memberikan rekomendasi bagi madrasah untuk meningkatkan tingkat kelulusan. Hasilnya menunjukkan akurasi sebesar 88,24%, recall 92,11%, dan precision 87,50%, menandakan kinerja baik algoritma Naïve Bayes dalam memprediksi kelulusan siswa.

Copyright © 2023 JSCR. All rights reserved.

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 mencenangkan program madrasah riset pada tahun 2013 (Agama et al., 2019). 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 diindonesia indeks yang menjadi syarat kelulusan bagi para siswa juga ikut naik (Mulyati et al., 2020).

Konsep dan definisi laporan Tugas Akhir menggunakan Teknik atau metode sistematis yang digunakan untuk mengklasifikasi data, metode yang akan digunakan adalah metode naïve bayes yang digunakan berdasarkan factor masalah mengenai hasil ujian nilai kelulusan siswa madrasah terutama yang belum memenuhi kriteria untuk melanjutkan ke jenjang kelas berikutnya. Tujuan dari penelitian ini untuk mengembangkan sebuah model klasifikasi menggunakan Teknik pembelajaran mesin untuk mengklasifikasikan siswa madrasah berdasarkan factor yang relevan. Definisi madrasah, Madrasah merupakan sebuah Lembaga Pendidikan agama yang berada di bawah naungan kementerian agama. 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.(Hidayatullah et al., 2022.).

Penelitian terdahulu yang dilakukan oleh penulis diantaranya oleh Edhy Poerwandono dengan judul Penerapan data mining Untuk penialian Kinerja Karyawan di Pt Riksa Dinar Djaya Menggunakan Metode Naïve Bayes Classification, masalah dalam penelitian ini adalah karena menurunnya kualitas pekerja dalam melakukan pekerjaannya dan masih terlihat banyak pekerja yang melanggar ketentuan SOP yang berlaku, tujuan dari penelitian ini agar karyawan bekerja dengan penuh semangat dan mematuhi SOP yang ada supaya Perusahaan lebih maju lagi, maka diterapkannya data mining dengan menggunakan metode naïve bayes untuk mendapatkan nilai yang kompeten dan relevan untuk dijadikan acuan sebagai penerapan kebijakan Perusahaan yang baru, hasil dari penelitian menggunakan algoritma naïve bayes adalah accuracy sebesar 99.44% dengan demikian penerapan naïve bayes dapat dijadikan alternatif untuk pengambilan keputusan dalam menilai kinerja karyawan di PT. Riksa Dinar Djaya. (Poerwandono & Perwitosari, 2023)

Berdasarkan permasalahan laporan tugas akhir dengan ini di sajikan data kelulusan nilai sekolah madrasah kelas 5 pada tabel 1.1 dengan terdapat 200 data dengan atribut jenis kelamin, asal, tahun masuk, nilai quran, nilai hadis, nilai akhlak, nilai akidah, nilai fikih, nilai Tarikh islam dan nilai Bahasa arab dan atribut label dengan class lulus dan belum lulus sebagai berikut:

Tabel 1. Dataset Kelulusan 1

No	Jenis Kelamin	Asal	Tahun Masuk	NQ	NH	NA	NAK	NF	NTI	NBA	Hasil
1	perempuan	Cirebon	2015	8	7.5	7.5	7.5	7.6	7.5	7.5	Belum lulus
2	Perempuan	Cirebon	2015	9	8.5	7.5	7.8	7.8	7.8	7.8	lulus
3	Perempuan	Cirebon	2015	9	9	7.5	8	7.8	7.5	8.5	lulus
4	Laki-laki	Cirebon	2015	8.8	9.5	8.5	8.3	7.5	7.5	7.5	lulus
5	Perempuan	Cirebon	2015	8	8	7	7	7.5	7.2	7.5	Belum lulus
6	Laki-laki	Cirebon	2015	7.3	7.5	8	8	7	7	7	Belum lulus
7	Laki-laki	Cirebon	2015	7.2	7.5	7	7.3	7	7	7	Belum lulus
8	Laki-laki	Cirebon	2015	9	7.5	7	8	8	8.5	7	lulus
9	Laki-laki	Cirebon	2015	8	9.5	8.5	8.5	7.2	7.5	7.5	lulus
10	perempuan	Cirebon	2015	8.8	9.5	8.5	8.3	7.5	7.5	7.5	lulus
...	...	...	...	...	...	...	...	...	...	...	...
200	perempuan	Majalengka	2019	9.5	9.6	9.1	9.5	8.9	8.9	8.8	lulus

Penjelsan Tabel 1. adalah data diatas merupakan data primer karena diambil secara langsung ke lokasi madrasah data tersebut berisi data sebagai berikut: pada kolom pertama berisi jenis kelamin, pada kolom kedua berisi asal, pada kolom ke tiga berisi Tahun Masuk, pada kolom ke keempat berisi nilai quran, pada kolom ke lima berisi nilai hadits, pada kolom ke enam berisi nilai akidah, pada kolom ke tujuh berisi nilai akhlak, pada kolom ke delapan berisi nilai Fikih dan pada kolom ke sembilan berisi nilai Tarikh islam, kolom ke sepuluh berisi nilai Bahasa arab kemudian ada atribut label pada kolom ke sebelas yang berisi *class* lulus dan belum lulus.

Permasalahan yang ditemukan dilapangan yang pertama data tersebut masih tercecer dan perlu mencarinya yang membutuhkan waktu cukup lama untuk mengumpulkan datanya tersebut, setelah data tersebut terkumpul maka perlu menginputkannya secara manual ke Microsoft excel dikarenakan data tersebut masih dalam bentuk buku yang juga membutuhkan waktu 1-2 hari untuk menginputkannya

Tujuan laporan tugas akhir ini adalah untuk mengembangkan sebuah model klasifikasi menggunakan Teknik pembelajaran mesin untuk mengklasifikasai berdasarkan kelulusan siswa dan meningkatkan minat belajar siswa dan mengetahui

hasil dari kelulusan para siswa apakah dinyatakan lulus atau belum lulus kemudian untuk siswa yang nilainya dibawah kriteria kelulusan atau belum lulus akan tetap lulus namun bersyarat apakah itu ada tugas tambahan atau program pengulangan agar mencapai standar kelulusan dan memberikan rekomendasi bagi madrasah untuk meningkatkan tingkat kelulusan siswanya.

Hasil dari laporan tugas akhir berikut didapatkan hasil yaitu tingkat accuracy sebesar 88,24%, recall sebesar 92,11% dan precision 87,50% dan roc sebesar 0.922 yang berarti dalam Laporan Tugas Akhir ini sudah cukup baik dalam mengklasifikasi antara siswa yang lulus dengan siswa yang belum lulus yang kemudian menjadi bahan acuan untuk para guru untuk meningkatkan prestasi siswanya.

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

METODE

Sumber Data

Dataset ini didapatkan dari sekolah madrasah nurul anwar yang berlokasi di Jln Pangeran Antasari Kec. Plumbon Kab. Cirebon 45155 dengan nama data kelulusan siswa kelas 5, data diambil pada tanggal 1 Desember 2023

Tabel 2. Sumber Data 1

No	Jenis Kelamin	Asal	Tahun Masuk	NQ	NH	NA	NAK	NF	NTI	NBA	Hasil
1	perempuan	Cirebon	2015	8	7.5	7.5	7.5	7.6	7.5	7.5	Belum lulus
2	Perempuan	Cirebon	2015	9	8.5	7.5	7.8	7.8	7.8	7.8	lulus
3	Perempuan	Cirebon	2015	9	9	7.5	8	7.8	7.5	8.5	lulus
4	Laki-laki	Cirebon	2015	8.8	9.5	8.5	8.3	7.5	7.5	7.5	lulus
5	Perempuan	Cirebon	2015	8	8	7	7	7.5	7.2	7.5	Belum lulus
6	Laki-laki	Cirebon	2015	7.3	7.5	8	8	7	7	7	Belum lulus
7	Laki-laki	Cirebon	2015	7.2	7.5	7	7.3	7	7	7	Belum lulus
8	Laki-laki	Cirebon	2015	9	7.5	7	8	8	8.5	7	lulus
9	Laki-laki	Cirebon	2015	8	9.5	8.5	8.5	7.2	7.5	7.5	lulus
10	perempuan	Cirebon	2015	8.8	9.5	8.5	8.3	7.5	7.5	7.5	lulus
...	...	...	...	...	...	...	...	...	...	...	...
200	perempuan	Majalengka	2019	9.5	9.6	9.1	9.5	8.9	8.9	8.8	lulus

Pada Tabel 1 merupakan dataset kelulusan kelas 5 madrasah dari tahun 2019 hingga tahun 2023 untuk datanya merupakan data primer karena diambil langsung ke madrasah nurul anwar. Untuk class yang belum lulus bukan berarti para siswa tidak lulus akan tetapi ada beberapa mata Pelajaran yang dinyatakan nilainya di bawah standar kelulusan atau lulus namun bersyarat apakah akan ada tugas tambahan atau adanya pertimbangan dari para guru atau bahkan tidak akan ada perbaikan sama sekali. Untuk nilai KKM sebagai berikut: Bahasa arab: 7.2, Fikih: 7.2, Tarikh Islam: 7.2, Al-quran: 7.4, Hadis: 7.3, akhlak: 7.1 dan akidah: 7.1 akan tetapi untuk nilai tulis tersebut belum sepenuhnya untuk kelulusan karena akan ada nilai tambah bukan hanya ujian tulis yaitu ujian praktek seperti shalat, hafalan surat al-quran dan lainnya untuk memberikan nilai tambah. Karena ijazah Madrasah ini diperlukan sebagai syarat untuk mendaftar ke smp negeri. Untuk yang saya ambil untuk laporan tugas akhir hanya ujian tulisnya saja.

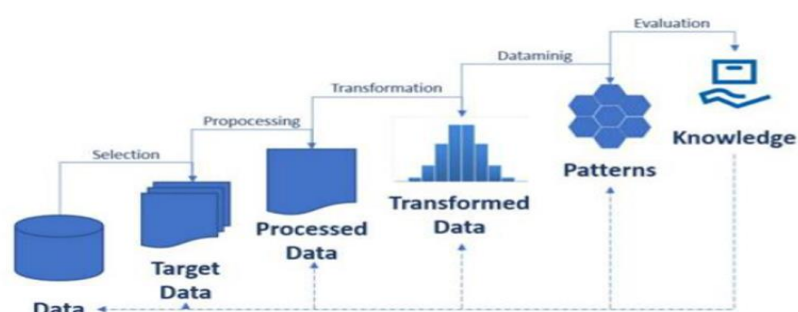
Teknik Pengumpulan Data

tahapan untuk pengumpulan data sendiri saya menggunakan wawancara kepada guru di madrasah terkait untuk mendapatkan data, Teknik wawancara saya gunakan karena saya tidak mengamati secara langsung ujian kelulusan siswa karena ujian telah selesai 3 minggu sebelumnya Adapun tahapannya sebagai berikut:

1. Wawancara: pada tanggal 1 Desember 2023 saya melakukan kunjungan ke madrasah Nurul anwar di JL. Pangeran Antasari untuk menemui wakil kepala sekolah yang sebelumnya sudah memiliki janji untuk datang di tanggal tersebut.
2. Wawancara: Pada hari yang sama saya meminta izin kepada wakil kepala sekolah di Madrasah tersebut guna meminta izin mendapatkan data untuk membuat laporan tugas akhir
3. Setelah meminta izin kemudian saya bertanya kepada wakil kepala madrasah, bertanya tentang visi dan misi madrasah, Sejarah berdirinya madrasah, struktur organisasi madrasah, tugas dan wewenang di madrasah tersebut dan mengenai data yang ingin saya ambil untuk membuat laporan tugas akhir.
4. Pengolahan Data: setelah mendapatkan izin kemudian pada esok hari saya menemui guru kelas 5 dan meminta izin untuk menganalisis data kelulusan dari tahun 2019 sampai 2023 dan saya mendapatkan data hasil ujian tulis tersebut dalam bentuk buku, untuk ujian prakteknya saya tidak mendapatkan karena datanya terlalu sedikit.
5. Pengumpulan data: setelah mendapatkan data saya melakukan analisis terhadap data apakah data tersebut sesuai dengan yang ingin saya gunakan untuk membuat laporan tugas akhir dari data tersebut didapatkan untuk tahun 2015 sebanyak 35 data dengan 21 lulus dan 14 belum lulus, tahun 2016 sebanyak 37 data dengan 20 lulus dan 17 belum lulus, tahun 2017 sebanyak 44 data dengan 26 lulus dan 18 belum lulus, tahun 2018 sebanyak 50 dengan 28 lulus dan 22 belum lulus tahun 2019 sebanyak 34 data dengan 19 lulus dan 15 belum lulus Dan data tersebut berisi 11 atribut termasuk label dengan kelas lulus dan belum lulus.
6. Penginputan dan perhitungan data: setelah data itu dianalisis dan sesuai dengan yang ingin saya gunakan selanjutnya saya melakukan penginputan data kedalam Microsoft excel untuk di jadikan dataset dan melakukan perhitungan dengan rapid miner dan didapatkan hasil yang baik.
7. Penyusunan Laporan: setelah selesai proses perhitungan, kemudian saya Menyusun laporan untuk membuat tugas akhir dan saya memberi judul "Implementasi data mining untuk mengklasifikasi hasil kelulusan siswa Madrasah menggunakan algoritma naïve bayes".

Teknik analisa data

Pada tahapan Analisa data menggunakan metode *knowledge discovery in database* atau KDD yang tahapannya selection, pre-processing, transformation, data mining dan evaluation pada Gambar 1.



Gambar 1. Metode Analisa Data 1

Tahapan penelitian dilakukan dengan menggunakan KDD (*knowledge Discovery in Database*) adapun langkanya sebagai berikut:

1. Data Selection

Merupakan sebuah proses menyeleksi data atau menyiapkan data yang akan dijadikan atribut/indicator/sebuah target dalam sebuah data yang besar.

2. Pre-Processing/Cleaning

Merupakan sebuah proses untuk memperbaiki kesalahan dalam data, memilah data yang ganda dan data yang dianggap kesalahan cetak.

3. 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.

4. Data Mining

Merupakan proses pencarian informasi atau pola yang menarik dan bermanfaat dengan menggunakan Teknik atau metode tertentu.

5. Evaluation

Merupakan sebuah proses untuk menampilkan hasil dan informasi dari proses data mining ke dalam sebuah bentuk yang mudah untuk dimengerti.

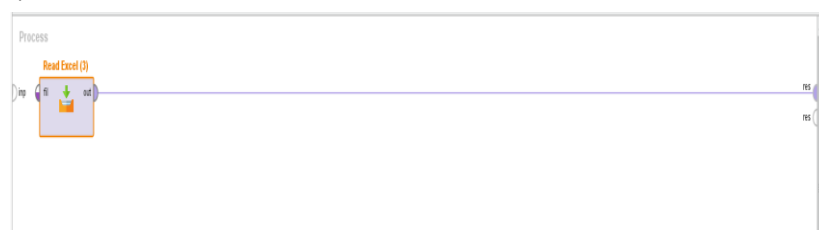
HASIL DAN PEMBAHASAN

Hasil

Untuk mendapatkan hasil pada lapoan tugas akhir ini menggunakan metode *knowledge Discovery in database* (KDD) dengan tahapan Data selection, Pre-Processing, Transformation, Data mining dan evaluasi pada laporan tugas akhir ini menggunakan algoritma klasifikasi yaitu *Naïve Bayes classification* untuk mendapatkan hasil klasifikasi berupa confusion matriks, recall, precision. Untuk tahapan KDD nya sebagai berikut:

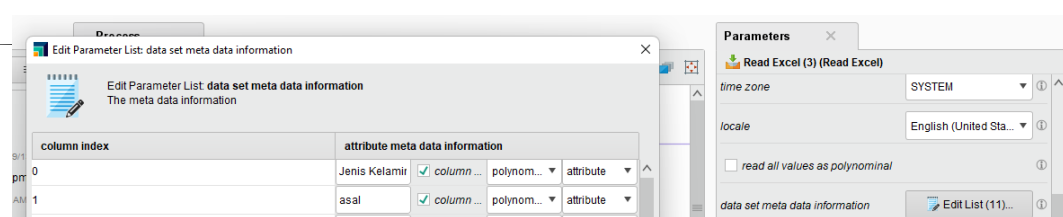
1. Data selection

Model yang digunakan dalam data selection menggunakan operator read excel yang berfungsi membaca dataset excel yang akan kita gunakan untuk model pada Gambar 2.



Gambar 2. Read Excel 1

Tahapan pertama membuat model dengan menggunakan operator read excel pada gambar 2 berikut dengan memasukan dataset kedalam operator read excel. Setelah memasukan data kemudian hubungkan ouput dengan result lalu klik tombol *run* yang ada diatas dan didapatkan hasil pada Gambar 3.



Gambar 3. Parameter read excel

Dari gambar 3 parameter read excel tidak ada yang dirubah, pada tahapan selection pada gambar 3 Atribut yang akan digunakan dalam penelitian ada 11 atribut dan salah satunya adalah label untuk atributnya adalah jenis kelamin, asal, Tahun Masuk, nilai Qur'an, nilai Hadits, nilai akidah, nilai fikih, nilai Tarikh islam dan nilai Bahasa arab dan Hasil sebagai label dengan kelas lulus dan Belum lulus. Untuk hasil dari tahapan ini dimuat ke dalam format pada Tabel 3.

Tabel 3. Data Selection 1

No	Jenis Kelamin	Asal	Tahun Masuk	NQ	NH	NA	NAK	NF	NTI	NBA	Hasil
1	perempuan	Cirebon	2015	8	7.5	7.5	7.5	7.6	7.5	7.5	Belum lulus
2	Perempuan	Cirebon	2015	9	8.5	7.5	7.8	7.8	7.8	7.8	lulus
3	Perempuan	Cirebon	2015	9	9	7.5	8	7.8	7.5	8.5	lulus
4	Laki-laki	Cirebon	2015	8.8	9.5	8.5	8.3	7.5	7.5	7.5	lulus
5	Perempuan	Cirebon	2015	8	8	7	7	7.5	7.2	7.5	Belum lulus
6	Laki-laki	Cirebon	2015	7.3	7.5	8	8	7	7	7	Belum lulus
7	Laki-laki	Cirebon	2015	7.2	7.5	7	7.3	7	7	7	Belum lulus
8	Laki-laki	Cirebon	2015	9	7.5	7	8	8	8.5	7	lulus
9	Laki-laki	Cirebon	2015	8	9.5	8.5	8.5	7.2	7.5	7.5	lulus
10	perempuan	Cirebon	2015	8.8	9.5	8.5	8.3	7.5	7.5	7.5	lulus
...	...	...	...	...	...	...	...	...	...	...	...
200	perempuan	Majalengka	2019	9.5	9.6	9.1	9.5	8.9	8.9	8.8	lulus

Hasil dari proses selection dalam bentuk tabel pada tabel 3. yang berjumlah 201 data dengan jumlah atribut sebelas atribut yang salah satunya adalah label.

2. Pre-processing/cleaning

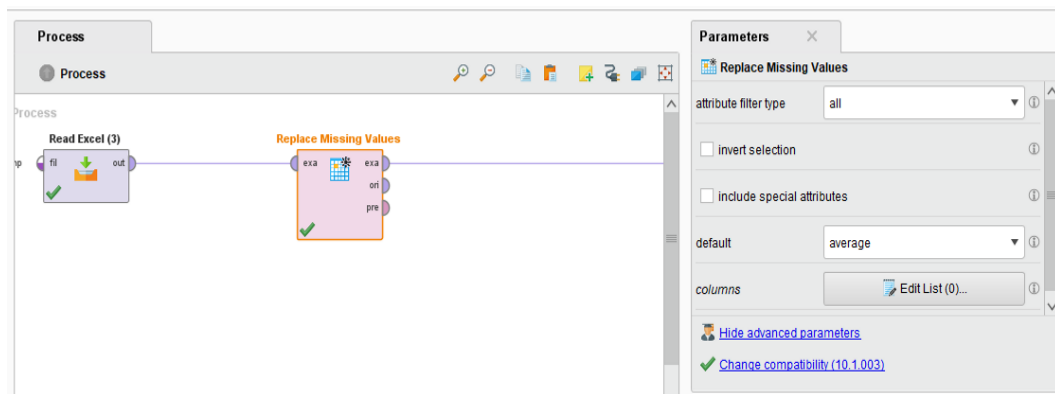
Operator yang akan digunakan dalam melakukan cleaning data adalah *replace missing value* yang merupakan operator yang digunakan untuk membersihkan data dan mencari data yang ganda ataupun data yang kosong.

Berikut model yang digunakan pada gambar 4.



Gambar 4. Replace Missing Value 1

Pada Gambar 4 operator *read excel* disambungkan dengan operator *replace missing value* dan dari *replace missing value* disambungkan ke *result* kemudian klik *run* dan dihasilkan hasil pada Gambar 5.



Gambar 5. Parameter Cleaning Data 1

Pada Gambar 5. parameter yang digunakan pada *attribute filter type* adalah *all* karena yang akan dibersihkan adalah semua data dan defaultnya adalah *average* yaitu rata-rata dari datanya dan untuk hasil dari melakukan *replace missing value* adalah pada Gambar 6.

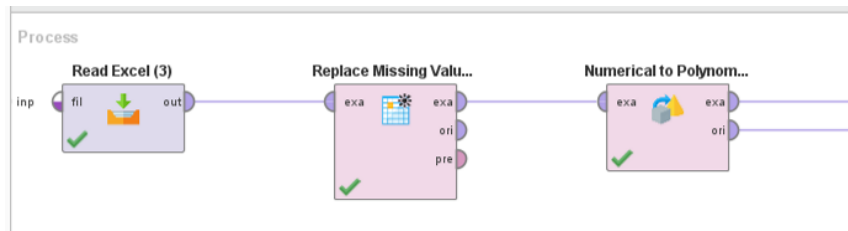
Row No.	Jenis Kelamin	asal	Tahun Masuk	nilai Quran	nilai hadis	nilai akidah	nilai akhlak	nilai fikih	nilai tarikh L...	nilai bahasa ...	Hasil
1	perempuan	cirebon	2015	8	7.500	7.500	7.500	7.600	7.500	7.500	belum lulus
2	perempuan	cirebon	2015	9	9	7.500	8	7.800	7.500	8.500	lulus
3	perempuan	cirebon	2015	8	8	7	7	7.500	7.200	7.500	belum lulus
4	laki-laki	cirebon	2015	7.300	7.500	8	8	7	7	7	belum lulus
5	laki-laki	cirebon	2015	9	7.500	7	8	8	8.500	7	lulus
6	laki-laki	cirebon	2015	8	9.500	8.500	8.500	7.200	7.500	7.500	lulus
7	perempuan	cirebon	2015	8.800	9.500	8.500	8.300	7.500	7.500	7.500	lulus
8	perempuan	cirebon	2015	9.500	9.500	9	9	8.500	8.500	9	lulus
9	perempuan	cirebon	2015	8.500	9	8	8	7.500	7.700	7.500	lulus
10	perempuan	cirebon	2015	8.500	8.800	7.500	7.500	7.500	7.500	8	lulus
11	perempuan	cirebon	2015	8	8.700	8	7.500	7.500	7.500	7.200	belum lulus
12	perempuan	cirebon	2015	9	8.500	7.500	7.800	7.800	7.800	7.800	lulus
13	laki-laki	cirebon	2015	8.800	9.500	8.500	8.300	7.500	7.500	7.500	lulus
14	laki-laki	cirebon	2015	7.200	7.500	7	7.300	7	7	7	belum lulus
15	perempuan	cirebon	2015	8	8.500	7.500	7.500	7.500	7.500	7.500	belum lulus

Gambar 6. Hasil Replace Missing Value 1

Dari hasil *replace missing value* pada Gambar 6 dari data tersebut yang berjumlah 200 data baik dari kolom atribut pertama hingga ke sebelas tidak ada data yang null dan missing itu berarti data tersebut dalam keadaan baik dan bisa di lanjutkan ke tahapan selanjutnya.

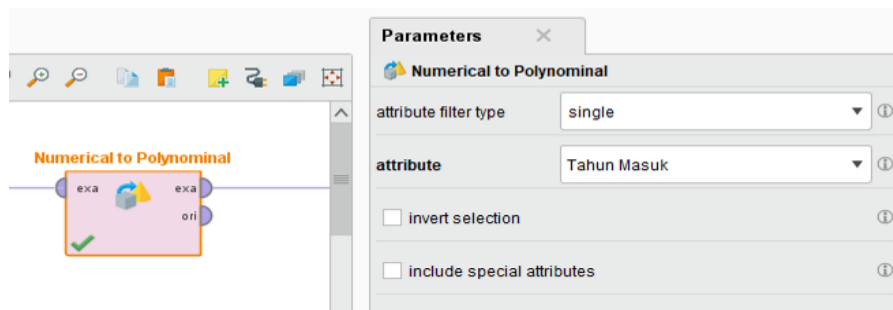
3. Transformation

Tahapan dalam melakukan transformation data dengan menggunakan *numerical to polynominal*, pada Operator *numerical to polynominal* berfungsi untuk merubah data numerik menjadi polynominal atau kategori adapun tahapan dalam *Rapid Miner* pada Gambar 7 sebagai berikut:



Gambar 7. Transformation 1

Pada Gambar 7 tahapan ini setelah melakukan cleaning data selanjutnya kita tambahkan operator baru yaitu *numerical to polynominal* kemudian sambungkan titik *example* dan *original* ke titik *result* kemudian klik *run*.



Gambar 8. Parameter transformation 1

Pada Gambar 8 parameter yang yang digunakan yaitu *attribute filter type* yang tadinya *all* di rubah menjadi *single* karena hanya satu atribut yang akan saya ubah dari *integer* ke *polynominal* yaitu atribut Tahun Masuk saja seperti pada Gambar 8.

Jadi dapat disimpulkan dalam model tersebut menghasilkan hasil pada gambar 9. sebagai berikut:

Format your columns.

☐ Replace errors with missing values ⓘ

	Jenis Kela... <i>polynominal</i>	asal <i>polynominal</i>	Tahun Masuk <i>polynominal</i>	nilai Quran <i>polynominal</i>	nilai hadis <i>polynominal</i>	nilai akidah <i>real</i>
1	perempuan	cirebon	2015	8.800	7.500	7.500
2	perempuan	cirebon	2015	8.000	7.500	7.500
3	perempuan	cirebon	2015	7.300	7.500	7.500
4	laki-laki	cirebon	2015	7.200	7.500	7.000
5	perempuan	cirebon	2015	9.000	9.500	8.500
6	laki-laki	cirebon	2015	8.800	9.500	9.000
7	laki-laki	cirebon	2015	8.000	8.500	7.500
8	laki-laki	cirebon	2015	8.800	7.500	7.500
9	laki-laki	cirebon	2015	9.500	7.500	7.500
10	perempuan	cirebon	2015	8.800	7.500	7.500
11	perempuan	cirebon	2015	8.800	7.500	7.500
12	perempuan	cirebon	2015	8.800	7.500	7.500

no problems.

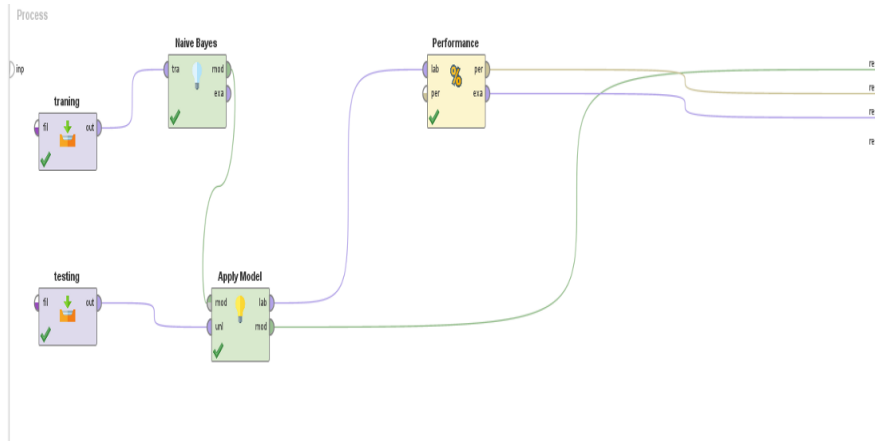
Previous Finish Cancel

Gambar 9. Hasil Transformation 1

Untuk hasilnya seperti gambar 9 untuk tahun masuk yang tadinya *integer* diubah ke *polynomial* karena untuk tahun sifatnya kategori karena untuk tahun masuk berisi lima kelas yaitu kelas 2015, 2016, 2017, 2018 dan 2019 dan juga agar hasil yang didapatkan ketahap selanjutnya menjadi lebih baik.

4. Data Mining

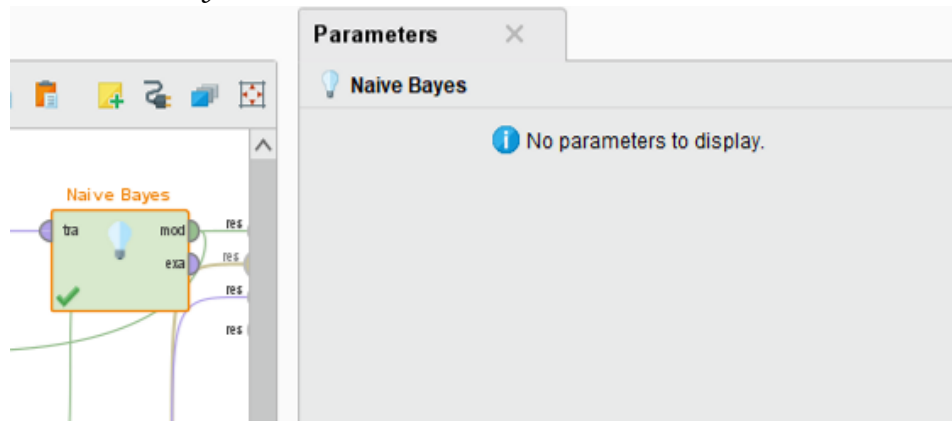
Pada model data mining menggunakan operator read excel 2 operator untuk data *traning* dan test, *naïve bayes*, dan ada tambahan operator yaitu *apply model* dan *performance* seperti pada gambar 10. Berikut:



Gambar 10. Model Data Mining 1

Pada gambar 10 Merubah nama *read excel* menjadi *traning* dan *testing* dan isikan sesuai dengan nama, kemudian data *traning* disambungkan ke *naïve bayes* dan data uji ke *apply model* dan *naïve bayes* disambungkan juga ke *apply model* setelah itu *apply model* disambung ke *performance* dan ke result lalu klik run. Untuk operator tambahannya sebagai berikut:

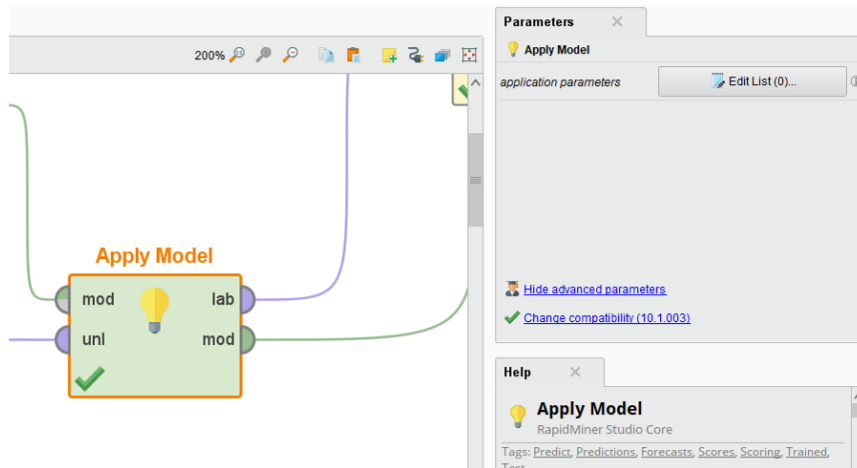
a. Operator *Naïve Bayes*



Gambar 11. Naive Bayes 1

Pada Gambar 11 merupakan operator *Naïve bayes* yang merupakan algoritma klasifikasi yang akan digunakan dalam laporan tugas akhir ini untuk parameter *naïve bayes* sendiri tidak ada yang dirubah.

b. Operator *apply model*



Gambar 12. performance 1

Pada gambar 12 Apply model digunakan untuk menerapkan sebuah data testing untuk mendapatkan sebuah klasifikasi atau prediksi pada data testing yang belum memiliki label pada model yang saya gunakan parameternya tidak ada yang diganti.

c. Operator *Performance*

Pada gambar 12 Operator performance digunakan untuk memberikan hasil evaluasi kinerja dan memberikan nilai secara otomatis. Pada parameters performance menggunakan use example weights atau contoh bobot, parameter sesuai dengan rapid miner tidak ada perubahan.

d. Evaluation

1. *Confusion Matrix*

Untuk hasil dari Rapid miner untuk confusion matrix dapat menghasilkan hasil pada gambar 13 sebagai berikut:

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

Gambar 13. Confusion matrix 1

Pada gambar 13 dapat diuraikan sebagai berikut:

true positive: 35

false positive: 5

true negative: 25

false negative: 3

2. *Acuray*

Akurasi yang didapatkan dalam Laporan Tugas Akhir menggunakan rapid nimer adalah pada gambar 14 berikut:

accuracy: 88.24%

	true lulus	true belum lulus	class precision
pred. lulus	35	5	87.50%
pred. belum lulus	3	25	89.29%
class recall	92.11%	83.33%	

Gambar 14. Acuracy 1

Untuk accuracy yang didapatkan dalam aplikasi rapidminer adalah 88,24%. Pada gambar 4.14 Dapat disimpulkan perhitungannya seperti di bawah ini:

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

Maka didapatkan hasil dari perhitungannya adalah 88,24%.

3. Precision

Hasil dari precision dalam Laporan Tugas Akhir menggunakan rapid miner pada gambar 15 berikut:

precision: 87.50% (positive class: lulus)

	true belum lulus	true lulus	class precision
pred. belum lulus	25	3	89.29%
pred. lulus	5	35	87.50%
class recall	83.33%	92.11%	

Dapat disimpulkan untuk hasil *precision* pada gambar 4.15 Adalah sebagai berikut

$$Precision = \frac{35}{(35+5)} = 0.875 \times 100\% = 87,50\%$$

Maka didapatkan hasil dari perhitungan adalah 87,50%

4. Recall

Hasil dari recall dalam Laporan Tugas Akhir menggunakan menggunakan rapid miner pada gambar 16. Sebagai berikut:

recall: 92.11% (positive class: lulus)

	true belum lulus	true lulus	class precision
pred. belum lulus	25	3	89.29%
pred. lulus	5	35	87.50%
class recall	83.33%	92.11%	

Gambar 16. Recall 1

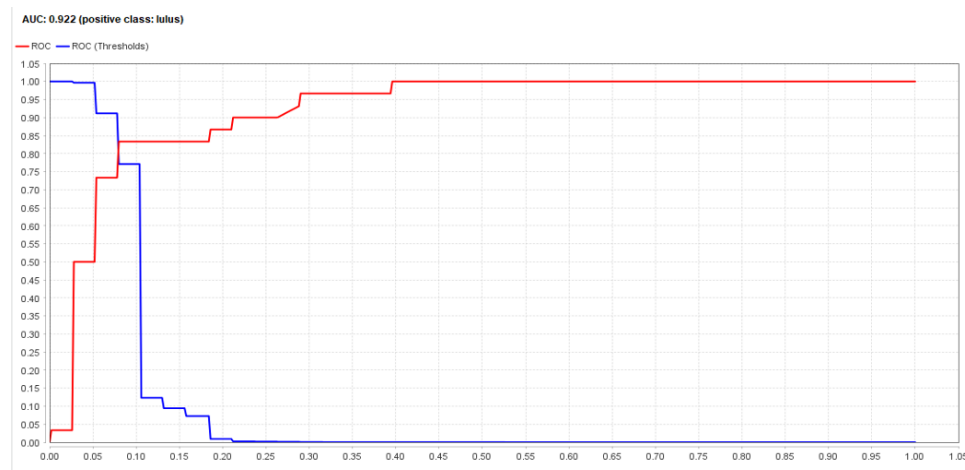
Dapat disimpulkan untuk perhitungan recall pada gambar 4.16 Adalah sebagai berikut:

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

Maka didapatkan hasil dari perhitungan adalah 92,11 %

5. ROC

Di dalam rapid miner ada yang Namanya kurva AOC digambarkan pada gambar 17 berikut:



Gambar 17. ROC 1

Untuk kurva ROC pada gambar 4.17 didapatkan dalam rapid miner adalah sebesar 0.922 excelent yang berarti algoritma naïve bayes ini menghasilkan hasil yang akurat.

6. Probabilitas prior

Hasil dari perhitungan menggunakan rapid miner dihasilkan dari data traning pada gambar 18. Berikut:

SimpleDistribution

Distribution model for label attribute Hasil

Class belum lulus (0.424)
10 distributions

Class lulus (0.576)
10 distributions

Gambar 18. Probabilitas Prior 1

Dapat disimpulkan dari perhitungan data traning pada gambar 4.18 Adalah sebagai berikut:

Lulus: 76/132 data = 0.576

Belum Lulus 56/132 data = 0.424

Dari perhitungan diatas 132 merupakan data traning yang digunakan dalam perhitungan naïve bayes.

7. Probabilitas jenis kelamin, asal dan tahun masuk berdasarkan class lulus dan Belum Lulus

Untuk probabilitas jenis kelamin dalam format Tabel 4. berikut:

Tabel 4. Probabilitas Jenis Kelamin 1

	Probabilitas jenis kelamin	
	Lulus	Belum lulus
Laki-laki	0.539	0.589
perempuan	0.4605	0.411

Pada Tabel 4. untuk probabilitas jenis kelamin laki laki dihitung berdasarkan jumlah laki-laki lulus dengan jumlah seluruh lulus begitu pun Perempuan rumus yang digunakan sama.

Tabel 5. Probabilitas Asal 1

	Probabilitas asal	
	Lulus	Belum Lulus
Bandung	0.013157895	0
Cirebon	0.973684211	0.892857143
Indramayu	0	0.035714286
Majalengka	0	0
Kuningan	0	0.017857143
Grobogan	0	0.017857143
Tangerang	0	0.035714286

Berdasarkan probabilitas pada Tabel 5 asal daerah dihitung berdasarkan asal Cirebon lulus dibagi total lulus begitupun pada semua data asal yang ada pada Tabel 6.

Tabel 6. Probabilitas Tahun Masuk 1

	Probabilitas Tahun Masuk	
	Lulus	Belum Lulus
2015	0.18421053	0.14285714
2016	0.19736842	0.21428571
2017	0.21052632	0.23214286
2018	0.23684211	0.26785714
2019	0.17105263	0.14285714

Untuk perhitungan probabilitas tahun masuk pada Tabel 6 kurang lebih sama dengan yang lainnya pada perhitungan ini menghasilkan hasil disajikan dalam bentuk tabel dalam perhitungan Tabel 4, Tabel 5 dan Tabel 6 menggunakan Microsoft excel dengan rumus =countifs ()/countif ()

8. Standar deviasi dan mean

Pada Tabel 6 dan 7 merupakan hasil perhitungan dari standar mean dan standar deviasi dengan menggunakan perhitungan rapid miner dan microsoft excel dan hasilnya sebagai berikut:

Attribute	Parameter	belum lulus	lulus
Tahun Masuk	value=unknown	0.000	0.000
nilai Quran	mean	7.696	8.532
nilai Quran	standard deviation	0.547	0.588
nilai hadis	mean	7.645	8.889
nilai hadis	standard deviation	0.540	0.648
nilai akidah	mean	7.479	8.366
nilai akidah	standard deviation	0.475	0.612
nilai akhlak	mean	7.516	8.437
nilai akhlak	standard deviation	0.556	0.605
nilai fikih	mean	7.379	8.004
nilai fikih	standard deviation	0.360	0.593
nilai tarikh islam	mean	7.304	7.987
nilai tarikh islam	standard deviation	0.342	0.564
nilai bahasa arab	mean	7.207	7.854
nilai bahasa arab	standard deviation	0.331	0.559

Gambar 19. mean dan Standar Deviasi 1

Hasil menggunakan rapid miner lebih akurat dibandingkan menggunakan microsoft excel pada Tabel 7 dan 8 Berikut:

Tabel 7. Mean 1

	nq	nh	nak	nakh	nf	Nti	Nba
L	8.531578947	8.889473684	8.365789474	8.436842105	8.003947368	7.986842105	7.853947368
BL	7.696428571	7.644642857	7.478571429	7.516071429	7.378571429	7.303571429	7.207142857

Pada Tabel 8. disajikan tabel mean yang merupakan rata rata dengan rumus total (penjumlahan)/total data

Tabel 8. Standar Deviasi 1

	nq	nh	nak	nakh	nf	Nti	Nba
L	0.587925285	0.647575776	0.611566869	0.605275057	0.593394931	0.563877553	0.55891342
BL	0.546713914	0.539669812	0.474711122	0.555559271	0.359653513	0.341634726	0.331309876

Untuk standar deviasi pada Tabel 8 untuk mendapatkan standar deviasi maka kita perlu menghitung meannya terlebih dahulu agar bisa menghitung standar deviasi.

9. Hasil lulus dan belum lulus

Pada Gambar 20 Merupakan hasil dari perhitungan menggunakan rapid miner:

Open in  Turbo Prep  Auto Model

Filter (68 / 68 examples): all

Row No.	Hasil	prediction(H...	confidence(L...	confidence(L...	Jenis Kelamin	asal	Tahun Masuk	nilai Quran	nilai hadis	nilai akidah	nilai akhlak	n
1	lulus	lulus	0.017	0.983	perempuan	cirebon	2015	9	8.500	7.500	7.800	7
2	lulus	lulus	0.000	1.000	laki-laki	cirebon	2015	8.800	9.500	8.500	8.300	7
3	belum lulus	belum lulus	1.000	0.000	laki-laki	cirebon	2015	7.200	7.500	7	7.300	7
4	belum lulus	belum lulus	0.955	0.045	perempuan	cirebon	2015	8	8.500	7.500	7.500	7
5	belum lulus	belum lulus	1.000	0.000	laki-laki	cirebon	2015	8	7	7.500	7	7
6	lulus	lulus	0.000	1.000	perempuan	cirebon	2015	8.500	9.500	8.200	8.500	7
7	lulus	lulus	0.000	1.000	perempuan	cirebon	2015	9	9.500	8	8.500	8
8	belum lulus	belum lulus	1.000	0.000	laki-laki	cirebon	2015	8	7	7.500	7	7
9	belum lulus	belum lulus	1.000	0.000	laki-laki	cirebon	2015	7	7	7.300	7	7
10	lulus	lulus	0.000	1.000	laki-laki	cirebon	2015	9	9.500	9	8.500	8
11	lulus	lulus	0.000	1.000	laki-laki	cirebon	2015	8.800	9.500	8.500	8.300	7
12	belum lulus	belum lulus	1.000	0.000	perempuan	cirebon	2015	8	7.500	7	7	7
13	lulus	lulus	0.000	1.000	laki-laki	cirebon	2015	9	9.500	8	8.500	8
14	lulus	lulus	0.000	1.000	laki-laki	cirebon	2016	9	9.100	7.500	9	8

ExampleSet (68 examples, 4 special attributes, 10 regular attributes)

Gambar 20. Hasil klasifikasi 1

Berdasarkan hasil pada gambar 20 diambil sekitar 68 data uji yaitu sebagai berikut untuk tahun 2015 diambil 13 data uji dari total 35 data dan tahun 2016 sekitar 10 data uji dari total 37 data, tahun 2017 sekitar 15 data uji dari 44 data, tahun 2018 sekitar 17 data uji dari total 51 data dan 2019 sekitar 13 data uji dari sekitar 34 data untuk hasilnya sebagai berikut untuk tahun 2015 dari 13 data uji terprediksi secara benar semua, untuk tahun 2016 dari 10 data terprediksi benar semua, untuk tahun 2017 dari 15 data ada 2 data belum lulus terprediksi lulus dan 2 data lulus terprediksi belum lulus, tahun 2018 dari 17 data 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.

Pembahasan

Pada Laporan Tugas Akhir yang telah dilakukan menghasilkan pola informasi dan pengetahuan baru, penggunaan data mining untuk mengklasifikasi hasil nilai kelas 5 madrasah. Menggunakan pola data mining yaitu: pola data training dan testing untuk mencari probabilitas setiap atributnya untuk mendapatkan sebuah informasi baru.

Nilai *accuracy* yang didapatkan dalam perhitungan menggunakan rapid miner didapatkan hasil yaitu 88,24%. Untuk *precision* dengan kelas lulus sebesar 87.50% dan untuk *recall* pada kelas lulus sebesar 92,11 % membuktikan bahwa metode ini cukup baik dalam melakukan klasifikasi.

Adapun atribut yang digunakan dalam mempengaruhi apakah hasil ujian tersebut dinyatakan lulus atau belum lulus adalah sebagai berikut: Jenis kelamin, asal, tahun masuk, nilai Quran, nilai hadits, nilai akidah, nilai akhlak, nilai fikih, nilai Tarikh islam dan nilai Bahasa arab.

SIMPULAN

Berdasarkan rumusan masalah pada bab 1 dapat disimpulkan untuk hasilnya adalah sebagai berikut.

1. Laporan Tugas Akhir yang dilakukan menghasilkan sebuah pola informasi dan pengetahuan baru. Pada Laporan Tugas Akhir ini data mining dan naïve bayes mampu mengklasifikasi hasil dengan sangat baik dengan menggunakan metode *knowledge discovery in database* dengan tahapan data selection, pre-processing, transformation, data mining dan evaluation yang nantinya akan menghasilkan informasi berdasarkan dataset kelulusan.
2. Berdasarkan dari dataset kelulusan dengan pembagain 132 data training dan 68 data testing dapat diperoleh hasil yaitu tingkat akurasi sebesar 88,24%, recall sebesar 92,11 % dan precision sebesar 87,50% dan roc sebesar 0.922 karena *accuracy* melebihi 50 % ini berarti kinerja naïve bayes dalam Laporan Tugas Akhir ini sudah cukup baik.

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 training 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

- Agama, P. P., Keagamaan, D., Litbang, B., Diklat, D., & Agama, K. (2019). 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. 17(3), 238-255. <http://jurnaledukasikemenag.org>
- Alfarisi, S. (2020). Analisis Pengembangan Komponen Kurikulum Pendidikan Islam di Madrasah Diniyah. *Rayah Al-Islam*, 4(02), 347-367. <https://doi.org/10.37274/rais.v4i02.346>
- Apriyani, H. (2020). Perbandingan Metode Naïve Bayes dan Support Vector Machine Dalam Klasifikasi Penyakit Diabetes Melitus. In *Journal of Information Technology Ampera* (Vol. 1, Issue 3). <https://journal-computing.org/index.php/journal-ita/index>
- Arfanda, I., Ramdhan, W., Yusda, R. A., & Artikel, H. (2021). Naive Bayes Dalam Menentukan Penerima Bantuan Langsung Tunai. 1(1). <https://doi.org/10.47709/briliance.vxix.xxxx>
- Chairiyah Madrasah Aliyah Negeri, Y. (2021). SEJARAH PERKEMBANGAN SISTEM PENDIDIKAN MADRASAH SEBAGAI LEMBAGA PENDIDIKAN ISLAM. In *Jurnal Pendidikan Islam* (Vol. 2, Issue 1).
- Hidayatullah, A. T., Islam, U., Hasan, Z., Probolinggo, G., Mahalli, J., & Yaqin, A. A. (2022). SEJARAH LEMBAGA PENDIDIKAN ISLAM (MADRASAH) DAN PERKEMBANGANNYA DI INDONESIA. *Jurnal Mahasiswa*, 4(3).

- Hoddin, M. S. (2020). Dinamika Politik Pendidikan... Dinamika Politik Pendidikan Islam Di Indonesia; Studi Kebijakan Pendidikan Islam Pada Masa Pra-Kemerdekaan Hingga Reformasi. In Jurnal Ilmiah Iqra (Vol. 14).
- Homepage, J., Musfiroh, D., Khaira, U., Eko, P., Utomo, P., Suratno, T., Studi, P., Informasi, S., Sains, F., & Teknologi, D. (2021). 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. 1, 24-33.
- Indra Borman, R., Ahmad, I., & Rahmanto, Y. (2022). Klasifikasi Citra Tanaman Perdu Liar Berkhasiat Obat Menggunakan Jaringan Syaraf Tiruan Radial Basis Function. Bulletin of Informatics and Data Science, 1(1). <https://ejurnal.pdsi.or.id/index.php/bids/index>
- Juliawati, F., Buaton, R., Saragih, R., & Kaputama, S. (2023). Pengelompokan Data Mining Penerimaan Bantuan Pangan Non Tunai (BPNT) Menggunakan Metode Clustering (Studi Kasus: Kantor Desa Payabakung Hamparan Perak). In Journal of Computer Science and Information Technology E-ISSN (Vol. 3, Issue 2).
- Kristiawan, K., & Widjaja, A. (2021). Perbandingan Algoritma Machine Learning dalam Menilai Sebuah Lokasi Toko Ritel. Jurnal Teknik Informatika Dan Sistem Informasi, 7(1). <https://doi.org/10.28932/jutisi.v7i1.3182>
- Kudus, U. M., Ganesha, J., & Kudus, P. (n.d.). Fida Maisa Hana.
- Mariana, D., & Mahrus Helmi, A. (2022). Madrasah Sebagai Lembaga Pendidikan Di Indonesia.
- Mulyati, S., Maulana Husein, S., & Kunci, K. (2020). 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. 65-73.
- Nasrullah, A. H. (2021). IMPLEMENTASI ALGORITMA DECISION TREE UNTUK KLASIFIKASI PRODUK LARIS. 7(2). <http://ejournal.fikom-unasman.ac.id>
- Nicodemus Turnip, H., & Fahmi, H. (2021). 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, 4(2), 36-41.
- Poerwandono, E., & Perwitosari, J. (2023). 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, 5(1), |pp. <https://doi.org/10.55338/saintek.v5i1.1416>
- Savitri, I., 1□, H., Defit, S., & Nurcahyo, G. W. (2021). Jurnal Sistim Informasi dan Teknologi <https://jsisfotek.org/index.php> Simulasi dalam Optimalisasi

Pengadaan Barang menggunakan Metode K-Mean Clustering.
<https://doi.org/10.37034/jsisfotek.v3i3.79>

Tarigan, P. M. S., Hardinata, J. T., Qurniawan, H., Safii, M., & Winanjaya, R. (2022). Implementasi Data Mining Menggunakan Algoritma Apriori Dalam Menentukan Persediaan Barang. *Jurnal Janitra Informatika Dan Sistem Informasi*, 2(1), 9–19. <https://doi.org/10.25008/janitra.v2i1.142>