

IMPLEMENTASI DATA MINING UNTUK MEMPREDIKSI KELULUSAN SISWA DENGAN METODE NAÏVE BAYES

Karolina Sinaga¹, Koko Handoko²

¹Alumni Program Studi Teknik Dan Komputer, Universitas Putera Batam

²Dosen Program Studi Teknik Dan Komputer, Universitas Putera Batam
email: karolinasinaga949@gmail.com

ABSTRACT

The problem of student grades is very important as a benchmark to see the level of student graduation. Effective value management with supporting applications that are very helpful in calculating accurate calculations. That way the use of naïve Bayes technology methods is able to answer problems in the information field of any integrated data. The purpose of this study was to determine the pass rate of students in the implementation of data mining for students who passed and did not pass, based on the final school examination scores (UAS), national test scores (UN), final scores (NA) for the last 3 years, 2017 to 2019. The data study method used in this research is by observation and interviews with resource persons from SMK Putra Jaya School Batam. The auxiliary application in the study used was WEKA to calculate student graduation results. This study used 70 student data as a test of value to be processed and produced a total of 210 data with 167 students who passed and 43 students who did not pass.

Keywords: *Data Mining, Naïve Bayes, WEKA.*

PENDAHULUAN

Pendidikan merupakan suatu sistem dimana terdiri dari banyaknya komponen yang saling berkaitan dan sangat erat namun mempunyai tujuan besar yang sama yaitu menghasilkan sumber daya yang berkualitas. Pendidikan memberikan pengaruh terhadap kualitas sumber daya manusia (SDM) yang sangat menentukan nasib bangsa. Sekolah menengah kejuruan (SMK) Putera Jaya School merupakan SMK kesehatan dan teknologi. Salah satunya

Jurusan Farmasi dan Keperawatan. Sampai tahun 2020 ini jurusan ini sudah mempunyai 1358 siswa dengan jumlah lulusan 744 siswa. Jumlah tersebut tidak kecil lagi bagi jurusan yang sudah berdiri sejak 8 tahun ini. Sekolah menengah kejuruan (SMK) Putera Jaya School tentu memiliki segudang data mengenai siswanya yang meliputi: data induk siswa dan data kelulusan siswa. Bertumpuknya data-data ini belum digunakan secara optimal padahal tumpukan-tumpukan



data tersebut dapat dijadikan suatu informasi yang sangat berharga.

Data dalam sekolah tersebut semakin lama akan semakin menumpuk dan disimpan tiap harinya tiap bulannya dan tiap tahunnya sehingga pihak sekolah sulit memprediksi tingkat kelulusan siswa. Maka dibutuhkan suatu teknik untuk membantu menemukan informasi-informasi berharga tersebut.

Data *mining* adalah proses menemukan pola atau informasi menarik dalam data yang dipilih menggunakan teknik atau metode. Metode *Naïve Bayes* merupakan bentuk dari suatu algoritma dalam teknik klasifikasi.

KAJIAN TEORI

2.1 Knowledge Discovery in Database (KDD)

Data *mining* juga dikatakan *Knowledge Discovery in Database* (KDD) membentuk aktifitas dengan cakupan pengumpulan, pemakaian historis yang akan menghasilkan kesesuaian, pola atau hubungan pada *set data* yang berukuran besar. *Knowledge Discovery In Database* (KDD) adalah sebuah langkah dalam mencatat bentuk data yang bersifat baru, sah dan dapat bermanfaat *Knowledge Discovery in Database* (KDD) merupakan proses untuk menemukan informasi yang berguna dalam *database*. (Elisa & Ervina, 2018).

2.2 Data Mining

Data *mining* adalah sebuah proses dimana satu atau lebih metode pembelajaran sistem komputer yang diberikan tugas dalam mengambil dan menganalisa pengetahuan secara langsung bahkan sekumpulan proses dalam menggali nilai tambah dari sekumpulan data yang tidak diketahui secara manual. (Elisa & Ervina, 2018).

Salah satu metode dalam teknik data mining adalah Metode *Naïve Bayes*. Metode *Naïve Bayes* atau Teorema keputusan bayes merupakan pendalaman statistic yang akurat untuk persepsi bentuk (*Pattern Recognition*). Persepsi ini dilandasi oleh kuantifikasi *trade-off antara* bermacam ketetapan klasifikasi dengan menggunakan probabilitas dan biaya yang dihasilkan pada ketetapan tersebut. (Diana Laily Fithri, 2016)

2.3 Penelitian Terdahulu

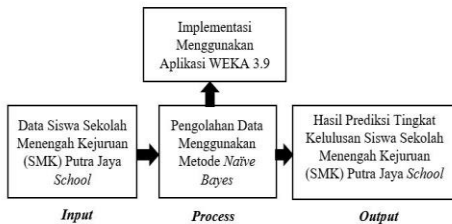
Penelitian terdahulu adalah tumpuan dari suatu penelitian.

1. Penelitian Rolando Marbun pada jurnal tahun 2019 yang berjudul **“Implementasi Data Mining Untuk Memprediksi Kelulusan Mahasiswa Menggunakan Algoritma *Naïve Bayes Classifier* Studi Kasus: Poltekkes Kemenkes Ri Medan”**. Pada penelitian ini, prediksi kelulusan mahasiswa pada Poltekkes Kemenkes Ri Medan dapat dilihat dari keputusan untuk mengatasi belajarnya. Untuk meringgankan bagian prodi untuk penemuan data mahasiswa yang selesai tepat waktu maupun tidak tepat waktu. Penelitian ini menggunakan metode *Naïve Bayes Classifier* yang merupakan sebuah pengklasifikasian probabilitas
2. Penelitian Yohakim Benedictus Samponu, Kusri pada jurnal tahun 2017 yang berjudul **“Optimasi Algoritma *Naïve Bayes* Menggunakan Metode *Cross Validation* Untuk Meningkatkan Akurasi Prediksi Tingkat Kelulusan Tepat Waktu”**. Penelitian ini melakukan prediksi tingkat kelulusan tepat waktu dengan melakukan analisis menggunakan teknik

klasifikasi data mining. Penulis melakukan serangkaian skenario percobaan yang mampu memberikan perbedaan untuk seberapa akurat yang ditemukan.

2.4 Kerangka Pemikiran

Kerangka pemikiran merupakan suatu pola yang menerangkan secara garis besar mengenai alur logika yang akan digunakan dalam perjalanan penelitian.



Gambar 1 Kerangka Pemikiran (Sumber: Data Peneliti, 2020)

Gambar 1 menjelaskan tentang alur dari kerangka pemikiran yaitu data siswa yang bersumber dari data nilai ujian siswa pada Sekolah Menengah Kejuruan (SMK) Putra Jaya School yang digunakan sebagai *input* dalam penelitian, kemudian diproses kedalam metode *Naive Bayes*, kemudian diuji menggunakan aplikasi WEKA 3.9 sehingga memperoleh *output* berupa hasil prediksi tingkat kelulusan siswa dari Sekolah Menengah Kejuruan (SMK) Putra Jaya School

METODE PENELITIAN

3.1 Desain Penelitian

Untuk melaksanakan penelitian, peneliti merancang desain penelitian yang menjelaskan tentang keseluruhan alur kegiatan saat penelitian.



Gambar 2. Desain Penelitian (Sumber: Data Peneliti, 2020)

Gambar 2 menjelaskan tentang alur dari desain penelitian dimana pengumpulan data dilakukan untuk memperoleh informasi yang diperlukan untuk mencapai tujuan penelitian. Pengambilan data dilakukan pada Sekolah Menengah Kejuruan (SMK) Putra Jaya School, dengan menggunakan teknik wawancara, observasi dan *study* pustaka mengenai materi-materi tentang data *mining* serta metode *naive bayes* yang diperoleh melalui buku-buku dan jurnal nasional maupun internasional. Penyeleksian Data Peneliti melakukan penyeleksian data dengan tujuan untuk memperoleh data untuk diolah kedalam penelitian. Data yang diseleksi yaitu berupa data-data siswa pada Sekolah Menengah Kejuruan (SMK) Putra Jaya School dari tahun 2017 hingga tahun 2019. Selanjutnya melakukan pengolahan data yang telah diseleksi, selanjutnya diolah kedalam metode *naive bayes* dari teknik data *mining*. Langkah terakhir yaitu melakukan pengujian menggunakan aplikasi WEKA versi 3.9 dalam memprediksi tingkat kelulusan siswa pada Sekolah Menengah Kejuruan

(SMK) Putra Jaya School, maka hasil Penelitian Peneliti menampilkan hasil penelitian yang telah diuji melalui perhitungan teknik data *mining* dengan metode *Naïve Bayes* serta diuji menggunakan aplikasi WEKA versi 3.9, yang menghasilkan prediksi tingkat kelulusan siswa pada Sekolah Menengah Kejuruan (SMK) Putra Jaya School.

3.2 Operasional Variabel

Operasional variabel yang digunakan dalam penelitian ini adalah nilai ujian siswa pada jurusan farmasi dan kesehatan di Sekolah Menengah Kejuruan (SMK) Putra Jaya School. Untuk mendukung penelitian ini, peneliti mengambil hasil nilai ujian siswa yaitu nilai ujian akhir sekolah (UAS), ujian nasional (UN) dan akhir (NA) setiap Berdasarkan perancangan sistem dari gambar 3 diatas dapat jelaskan bahwa proses awal yang dilakukan pada penelitian ini yaitu memperoleh data nilai ujian siswa berdasarkan nilai ujian akhir sekolah (UAS), ujian nasional (UN) dan akhir (NA) dari sekolah menengah kejuruan (SMK) Putra Jaya School, kemudian melakukan proses penyeleksian data sampai menghasilkan data yang layak untuk diproses ke tahapan selanjutnya. Setelah memperoleh hasil penyeleksian data maka data siswa tersebut selanjutnya diolah dalam teknik data *mining* menggunakan metode *Naïve Bayes*, selanjutnya akan diuji dengan

tahunnya selama 3 tahun terakhir yaitu dari tahun 2017 hingga tahun 2019.

3.3 Metode Perancangan Sistem

Metode perancangan sistem pada penelitian ini dapat dilihat sebagai berikut:



Gambar 3 Metode Perancangan Sistem (Sumber: Data Peneliti, 2020)

menggunakan aplikasi WEKA versi 3.9, serupa penegasan hasil yang didapatkan dari perhitungan dengan metode *naïve bayes* sesuai dengan hasil pengujian dengan aplikasi WEKA versi 3.9.

HASIL DAN PEMBAHASAN

4.1 Arsitektur Sistem dan Pembahasan

Data yang digunakan dalam penelitian ini bersumber dari data satu ruang dengan jurusan farmasi dan kesehatan. Nilai yang bersumber dari mata pelajaran matematika, bahasa Inggris dan bahasa Indonesia. Berikut data penelitian dari SMK Putra Jaya School dapat dilihat pada tabel berikut:

Tabel 1. Data Nilai Matematika Siswa

No	NAMA	NISN	Mata Pelajaran	UAS	UN	NA	Keterangan
1	Abdul Munif	17331	Matematika	75	62	65	Lulus
2	Achmad Affan	17332	Matematika	60	75	60	Lulus

3	Al Azizur Rahman	17333	Matematika	65	25	45	Tidak Lulus
4	Bagus Arjaka	17334	Matematika	78	76	78	Lulus
5	Bunga Dahlia	17335	Matematika	54	25	39	Tidak Lulus
6	Dewi Lailatul Ulfa	17336	Matematika	80	65	65	Lulus
7	Dwi Alfian Bastiansyah	17337	Matematika	65	38	51	Tidak Lulus
8	Dwi Ainul Haqiqy	17338	Matematika	65	38	51	Tidak Lulus
9	Ellis Marisa	17339	Matematika	65	79	65	Lulus
10	Endang Safitri	17340	Matematika	75	80	75	Lulus

(Sumber:Data Peneliti, 2020)

Tabel 1 menjelaskan tentang data yang diperoleh dari proses observasi dan wawancara bersama pihak sekolah dari SMK Putra Jaya *School* masih berupa data tidak jadi, sebab data dari atribut masih bergabung dengan data yang tidak penting dalam penelitian. Data sampel harus melalui proses pembersihan terlebih dahulu. Setelah proses seleksi data selesai maka data yang diperoleh yaitu dari kelas matematika, bahasa

inggris, dan bahasa indonesia yang akan dijadikan sampel dalam proses selanjutnya dalam penelitian ini.

Setelah proses seleksi data selesai, maka langkah selanjutnya adalah melakukan pembuangan bagian yang tidak penting dalam penelitian ini yaitu nomor urut dan NISN. Maka hasil pembersihan data seperti pada tabel berikut:

Tabel 2 Data *Training* Siswa Setelah Eliminasi

Nama	Mata Pelajaran	UAS	UN	NA	Keterangan
Abdul Munif	Matematika	75	62	65	Lulus
Achmad Affan	Matematika	60	75	60	Lulus
Al Azizur Rahman	Matematika	65	25	45	Tidak Lulus
Bagus Arjaka	Matematika	78	76	78	Lulus
Bunga Dahlia	Matematika	54	25	39	Tidak Lulus
Dewi Lailatul Ulfa	Matematika	80	65	65	Lulus
Dwi Alfian Bastiansyah	Matematika	65	38	51	Tidak Lulus
Dwi Ainul Haqiqy	Matematika	65	38	51	Tidak Lulus
Ellis Marisa	Matematika	65	79	65	Lulus
Endang Safitri	Matematika	75	80	75	Lulus

(Sumber:Data Peneliti, 2020)

Tabel 2 diatas akan dipergunakan sebagai data coba atau data *training* dalam proses

pembuktian dengan aplikasi WEKA versi 3.9. Setelah itu, proses selanjutnya

adalah merancang suatu data coba yang dipenuhi data prediksi yang akan terjadi waktu ini maupun yang akan terjadi dimasa yang akan datang. Berikut adalah

data training yang akan diuji kedalam metode naïve bayes kemudian diolah kedalam aplikasi WEKA.

Tabel 3 Data *Testing*

NAMA	MATA PELAJARAN	UAS	UN	NA	KETERANGAN
Abdul Munif	Matematika	75	62	65	?
Bunga Dahlia	Matematika	54	25	39	?
Zulfianto Putra	Bahasa Inggris	44	23	25	?
Zikry Mahbubi	Bahasa Inggris	80	76	80	?
Yoga Prasetyo	Bahasa Indonesia	74	85	78	?

(Sumber:Data Peneliti, 2020)

Table 3 akan dijadikan data coba yang akan dihitung dengan metode naïve bayes berjumlah 5 data siswa dari 70 data siswa yang di ambil. Data coba yang dimasukkan adalah berupa data prediksi nilai siswa kedepannya untuk

menentukan tingkat kelulusan yaitu lulus dan tidak lulus.

1. Percobaan 1

Percobaan pertama dilakukan untuk mencari nilai dari kolom "KETERANGAN". Data *testing* yang diambil dapat dilihat pada tabel berikut:

Tabel 4 Data *Testing* 1

NAMA	MATA PELAJARAN	UAS	UN	NA	KETERANGAN
ABDUL MUNIF	MATEMATIKA	75	62	65	?

(Sumber:Data Peneliti, 2020)

Dalam proses *training* data, data *testing* yang diuji menggunakan acuan dari data *training* yang sudah dibuat sebelumnya.

$P(\text{KETERANGAN} = \text{"LULUS"}) = 47/70$
 $P(\text{KETERANGAN} = \text{"TIDAK LULUS"}) = 23/70$
 $P(\text{UAS} = \text{"75"} \mid \text{KETERANGAN} = \text{"LULUS"}) = 7/70$
 $P(\text{UAS} = \text{"75"} \mid \text{KETERANGAN} = \text{"TIDAK LULUS"}) = 0/70$

$P(\text{UN} = \text{"62"} \mid \text{KETERANGAN} = \text{"LULUS"}) = 6/70$
 $P(\text{UN} = \text{"62"} \mid \text{KETERANGAN} = \text{"TIDAK LULUS"}) = 0/70$
 $P(\text{NA} = \text{"65"} \mid \text{KETERANGAN} = \text{"LULUS"}) = 19/70$
 $P(\text{NA} = \text{"65"} \mid \text{KETERANGAN} = \text{"TIDAK LULUS"}) = 0/70$
 $P(\text{NAMA SISWA} = \text{"ABDUL MUNIF"} \mid \text{KETERANGAN} = \text{"LULUS"}) = 1/70$

$$P(\text{NAMA SISWA} = \text{"ABDUL MUNIF"} \mid \text{KETERANGAN} = \text{"TIDAK LULUS"}) = 0/70$$

$$P(\text{MATA PELAJARAN} = \text{"MATEMATIKA"} \mid \text{KETERANGAN} = \text{"LULUS"}) = 47/70$$

$$P(\text{MATA PELAJARAN} = \text{"MATEMATIKA"} \mid \text{KETERANGAN} = \text{"TIDAK LULUS"}) = 23/70$$

Maka perhitungan selanjutnya adalah sebagai berikut:

$$P(X \mid \text{KETERANGAN} = \text{"LULUS"}) = P(\text{UAS} = \text{"75"}, \text{UN} = \text{"62"}, \text{NA} = \text{"65"}, \text{NAMA SISWA} = \text{"ABDUL MUNIF"}, \text{MATA PELAJARAN} = \text{"MATEMATIKA"} \mid \text{KETERANGAN} = \text{"LULUS"})$$

$$P(X \mid \text{KETERANGAN} = \text{"LULUS"}) = 7/70 * 6/70 * 19/70 * 1/70 * 47/70$$

$$P(X \mid \text{KETERANGAN} = \text{"LULUS"}) = 2,1639079e^{-5}$$

$$P(X \mid \text{KETERANGAN} = \text{"TIDAK LULUS"}) = P(\text{UAS} = \text{"75"}, \text{UN} = \text{"62"}, \text{NA} = \text{"65"}, \text{NAMA SISWA} = \text{"ABDUL MUNIF"}, \text{MATA PELAJARAN} = \text{"MATEMATIKA"} \mid \text{KETERANGAN} = \text{"TIDAK LULUS"})$$

$$P(X \mid \text{KETERANGAN} = \text{"TIDAK LULUS"}) = 0/70 * 0/70 * 0/70 * 0/70 * 23/70$$

$$P(X \mid \text{KETERANGAN} = \text{"TIDAK LULUS"}) = 0$$

$$P(X|c) * (P|C)$$

$$P(X \mid \text{KETERANGAN} = \text{"LULUS"}) * P(\text{KETERANGAN} = \text{"LULUS"}) = 2,19482087e^{-5} * 0,671 = 1,4519822e^{-5}$$

$$P(X \mid \text{KETERANGAN} = \text{"TIDAK LULUS"}) * P(\text{KETERANGAN} = \text{"TIDAK LULUS"}) = 0 * 0,328 = 0$$

Maka hasil data percobaan 1 masuk kedalam kelas KETERANGAN " LULUS" atas nama ABDUL MUNIF.

Tabel 5 Hasil data *Testing* 1

NAMA	MATA PELAJARAN	UAS	UN	NA	KETERANGAN
ABDUL MUNIF	MATEMATIKA	75	62	65	LULUS

(Sumber:Data Peneliti, 2020)

2. Percobaan 2

Percobaan pertama dilakukan untuk mencari nilai dari kolom

"KETERANGAN". Data *testing* yang diambil dapat dilihat pada tabel berikut

Tabel 6 Data *Testing* 2

NAMA	MATA PELAJARAN	UAS	UN	NA	KETERANGAN
BUNGA DAHLIA	MATEMATIKA	54	25	39	?

(Sumber:Data Peneliti, 2020)

Dalam proses *training* data, data *testing* yang diuji menggunakan acuan dari data *training* yang sudah dibuat sebelumnya.

$$P(\text{KETERANGAN} = \text{"LULUS"}) = 47/70$$

$$P(\text{KETERANGAN} = \text{"TIDAK LULUS"}) = 23/70$$

$$P(\text{UAS} = \text{"54"} \mid \text{KETERANGAN} = \text{"LULUS"}) = 0/70$$

$$P(\text{UAS} = \text{"54"} \mid \text{KETERANGAN} = \text{"TIDAK LULUS"}) = 5/70$$

$$P(\text{UN} = \text{"25"} \mid \text{KETERANGAN} = \text{"LULUS"}) = 0/70$$

$$P(\text{UN} = \text{"25"} \mid \text{KETERANGAN} = \text{"TIDAK LULUS"}) = 13/70$$

$$P(\text{NA} = \text{"39"} \mid \text{KETERANGAN} = \text{"LULUS"}) = 0/70$$

$$P(\text{NA} = \text{"39"} \mid \text{KETERANGAN} = \text{"TIDAK LULUS"}) = 5/70$$

$$P(\text{NAMA SISWA} = \text{"BUNGA DAHLIA"} \mid \text{KETERANGAN} = \text{"LULUS"}) = 0/70$$

$$P(\text{NAMA SISWA} = \text{"BUNGA DAHLIA"} \mid \text{KETERANGAN} = \text{"TIDAK LULUS"}) = 1/70$$

$$P(\text{MATA PELAJARAN} = \text{"MATEMATIKA"} \mid \text{KETERANGAN} = \text{"LULUS"}) = 47/70$$

$$P(\text{MATA PELAJARAN} = \text{"MATEMATIKA"} \mid \text{KETERANGAN} = \text{"TIDAK LULUS"}) = 23/70$$

Maka perhitungan selanjutnya adalah sebagai berikut:

$$P(X \mid \text{KETERANGAN} = \text{"LULUS"}) =$$

$$P(\text{UAS} = \text{"54"}, \text{UN} = \text{"25"}, \text{NA} = \text{"39"}, \text{NAMA SISWA} = \text{"BUNGA DAHLIA"}, \text{MATA PELAJARAN} = \text{"MATEMATIKA"} \mid \text{KETERANGAN} = \text{"LULUS"})$$

$$P(X \mid \text{KETERANGAN} = \text{"LULUS"}) = 0/70 * 0/70 * 0/70 * 47/70$$

$$P(X \mid \text{KETERANGAN} = \text{"LULUS"}) = 0$$

$$P(X \mid \text{KETERANGAN} = \text{"TIDAK LULUS"}) = P(\text{UAS} = \text{"54"}, \text{UN} = \text{"25"}, \text{NA} = \text{"39"}, \text{NAMA SISWA} = \text{"BUNGA DAHLIA"}, \text{MATA PELAJARAN} = \text{"MATEMATIKA"} \mid \text{KETERANGAN} = \text{"TIDAK LULUS"})$$

$$P(X \mid \text{KETERANGAN} = \text{"TIDAK LULUS"}) = 5/70 * 13/70 * 5/70 * 1/70 * 23/70$$

$$P(X \mid \text{KETERANGAN} = \text{"TIDAK LULUS"}) = 4,28243032e^{-6}$$

$$P(X|c) * (P|C)$$

$$P(X \mid \text{KETERANGAN} = \text{"LULUS"}) * P(\text{KETERANGAN} = \text{"LULUS"}) = 0 * 0,671 = 0$$

$$P(X \mid \text{KETERANGAN} = \text{"TIDAK LULUS"}) * P(\text{KETERANGAN} = \text{"TIDAK LULUS"}) = 4,28243032e^{-6} * 0,328 = 1,4046374e^{-6}$$

Maka hasil data percobaan 1 masuk kedalam kelas KETERANGAN "TIDAK LULUS" atas nama BUNGA DAHLIA.

Tabel 7 Hasil data *Testing* 2

NAMA	MATA PELAJARAN	UAS	UN	NA	KETERANGAN
BUNGA DAHLIA	MATEMATIKA	54	25	39	TIDAK LULUS

(Sumber:Data Peneliti, 2020)

Setelah melakukan perhitungan manual dengan rumus dari metode *Naive Bayes* maka total keseluruhan data

menjadi 210 data. Kemudian dibedakan berapa yang lulus dn tidak lulus. Dapatkan dilihat pada tabel berikut ini:

Tabel 8 Hasil Akhir Dari Data Siswa

No	NAMA	MATA PELAJARAN	UAS	UN	NA	KETERANGAN
1	Abdul Munif	Matematika	75	62	65	Lulus
2	Achmad Affan	Matematika	60	75	60	Lulus
3	Al Azizur Rahman	Matematika	65	25	45	Tidak Lulus
4	Bagus Arjaka	Matematika	78	76	78	Lulus
5	Bunga Dahlia	Matematika	54	25	39	Tidak Lulus

(Sumber:Data Peneliti, 2020)

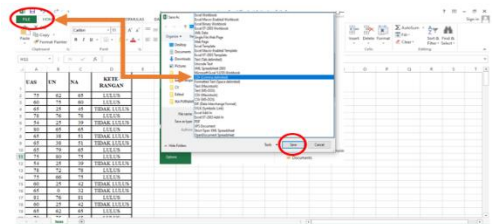
Tabel 8 menyatakan hasil pengujian yang telah dikerjakan dalam penelitian ini sebagai berikut:

1. Hasil prediksi tingkat kelulusan pada sekolah menengah kejuruan (SMK) Putra Jaya School di pengaruhi oleh 3 variabel yaitu variabel nilai ujian akhir sekolah (UAS), nilai ujian nasional (UN) dan nilai akhir (NA).
2. Jika dilihat dari data siswa pada sekolah menengah kejuruan (SMK) Putra Jaya School yang berjumlah 70 siswa yang telah mengikuti ujian akhir sekolah (UAS), ujian nasional (UN) dan memiliki nilai akhir (NA) pada jurusan farmasi dan kesehatan dengan tiga mata pelajaran yaitu matematika, bahasa inggris dan bahasa indonesia.
3. Total data siswa berjumlah 70 data dari tiga mata pelajaran maka menghasilkan 210 data *training* dengan 167 siswa yang lulus dan 43 siswa tidak lulus. Dimana terdapat 47 siswa yang dinyatakan lulus dan 23 siswa yang dinyatakan tidak lulus dari mata pelajaran matematika. Selanjutnya terdapat 50 siswa yang dinyatakan lulus dan 20 siswa yang dinyatakan tidak lulus dari mata pelajaran bahasa inggris. Sedangkan untuk mata pelajaran bahasa indonesia dinyatakan 70 siswa lulus dan 0 siswa yang tidak lulus.

4.2 Pengujian Data Menggunakan Aplikasi WEKA 3.9

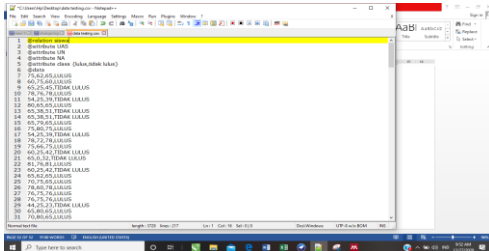
Setelah perhitungan manual selesai dilakukan, maka langkah selanjutnya yaitu pengujian terhadap kebenaran data-data dari proses perhitungan dengan metode naïve bayes menyamakan apakah hasil perhitungan tersebut sama atau tidak sama. Selanjutnya peneliti akan menguji keabsahan hasil perhitungan kedalam aplikasi WEKA versi 3.9. Dengan proses berikut:

1. mempersiapkan data dari semua variabel-variabel yang akan diproses kedalam aplikasi WEKA versi 3.9 *microsoft excel* dan *save as type CSV (Comma Delimited)*. Seperti gambar berikut:



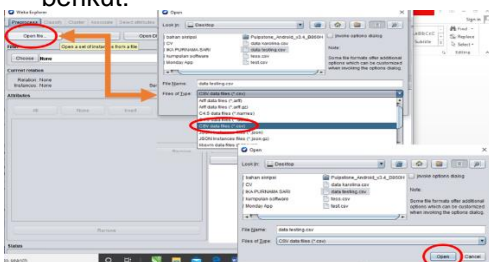
Gambar 4 Format Data *Microsoft Excel* (Sumber: Data Peneliti, 2020)

2. Selanjutnya buatlah sebuah *file .arff*, menggunakan *software notepad++* seperti gambar 4.2 berikut:



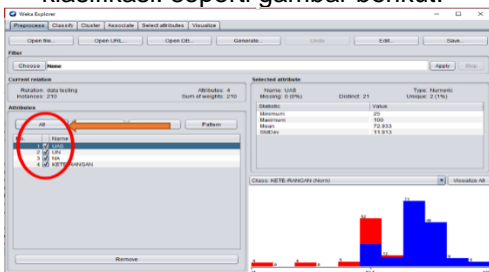
Gambar 5 Tampilan Data Pada *Notepad++*
(Sumber: Data Peneliti, 2020)

3. Pada aplikasi WEKA *explorer* langkah selanjutnya yaitu pilih *open file*, kemudian ambil data dari penyimpanan sebelumnya berupa file siswa dengan format *Files of Type CSV data files (*.csv)*, kemudian pilih *Open* . dapat dilihat pada gambar berikut:



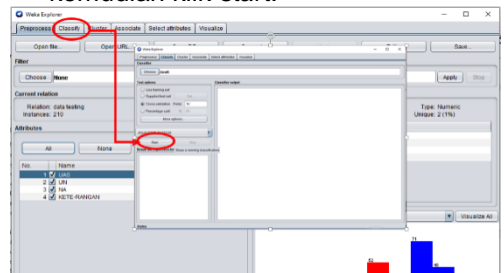
Gambar 6. Membuka Data Pada Menu *Explorer*
(Sumber: Data Peneliti, 2020)

4. Pilih variabel yang mempengaruhi penelitian, lalu klik *All* maka semua atribut dimasukkan ke dalam proses klasifikasi. seperti gambar berikut:



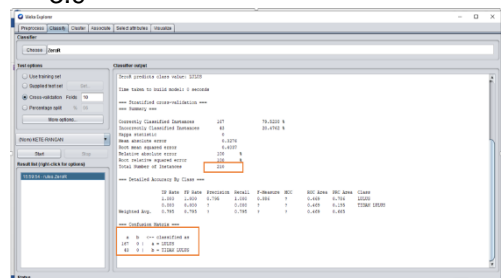
Gambar 7. Pemilihan Tabel Pada Aplikasi WEKA
(Sumber: Data Peneliti, 2020)

5. Jika semua variabel sudah dipilih dan dimasukkan, maka langkah selanjutnya yaitu klik menu *classify*, kemudian klik *start*.



Gambar 8. Pemilihan Model *Classify*
(Sumber: Data Peneliti, 2020)

6. Maka hasil akhir dari data siswa pada sekolah Putra Jaya School dengan menggunakan metode *naive bayes* serta di uji melalui aplikasi WEKA versi 3.9



Gambar 9. Hasil Pengujian Dengan Menggunakan Aplikasi WEKA

SIMPULAN

4.1 Simpulan

Berdasarkan hasil penelitian yang telah diperoleh, maka peneliti dapat merangkum kesimpulan bahwa:

1. Dengan adanya teknik data *mining* yang melibatkan 210 data siswa yang telah dihitung menggunakan metode

Naïve Bayes maka diperoleh hasil yaitu 167 siswa lulus dan 43 siswa tidak lulus, dari jurusan farmasi dan kesehatan pada Sekolah Menengah Kejuruan (SMK) Putra Jaya School dengan tiga mata pelajaran yaitu matematika, bahasa inggris dan bahasa indonesia yang dipengaruhi oleh tiga variabel yaitu nilai ujian akhir sekolah (UAS), nilai ujian nasional (UN) dan nilai akhir (NA) dengan hasil lulus dan tidak lulus.

2. Melalui perhitungan manual dengan menggunakan metode *naïve bayes* (Sumber: Data Peneliti, 2020)

kemudian diuji melalui Aplikasi WEKA dengan 5 data *testing*, sehingga memperoleh keakuratan antara perhitungan manual dan Aplikasi WEKA dalam memprediksi tingkat kelulusan siswa pada Sekolah Menengah Kejuruan (SMK) Putra Jaya School.

3. Mempercepat pengolahan nilai kelulusan siswa dan memiliki fungsi penting dalam membantu mendapatkan informasi nilai kelulusan siswa yang berguna dimasa yang akan datang.