

Penerapan Algoritma Naïve Bayes Pada Sistem Prediksi Tingkat Kelulusan Peserta Sertifikasi Microsoft Office Specialist (MOS)

Mochamad Farid Rifai¹; Hendra Jatnika²; Bowval Valentino³

^{1, 2, 3}Sekolah Tinggi Teknik PLN

¹m.farid@sttpln.ac.id

²h.jatnika@sttpln.ac.id

³bowval1431202@sttpln.ac.id

ABSTRACT

This research discusses prediction pass rates the certification microsoft office specialist 2013 version (word and excel) aimed to provide information concerning to pass rates and certification give alternative solutions to determine the program certification appropriate to chosen before test certification. Naive bayes used for the classification certification graduation where participants know what information pass and did not finish. Naive bayes is a classification with the probability and statistics to predict opportunities in the future based on the Provided before. In this study, system development CRISP-DM to use of the become more ordered and testing done with the BlackBox to test each function is on the application built. From the study, produce values probability of 0.001042 the accuracy of 99 %. These results, proving that naïve bayes method can be used to assist in a prediction graduation rates participants (word and excel), because it produces quite high accuracy. So participants were able to determine the certification program proper chosen before test certification.

Keywords: Prediction, Certification, Naïve Bayes, CRISP-DM, BlackBox

ABSTRAK

Penelitian ini membahas mengenai prediksi tingkat kelulusan peserta uji sertifikasi microsoft office specialist versi 2013 (word dan excel) bertujuan untuk memberikan informasi terkait tingkat kelulusan peserta sertifikasi dan memberikan alternatif solusi untuk menentukan program sertifikasi yang tepat untuk dipilih sebelum melakukan uji sertifikasi. Naïve bayes digunakan untuk klasifikasi kelulusan sertifikasi dimana peserta dapat mengetahui mengenai keterangan lulus dan tidak lulus. Naive Bayes merupakan pengklasifikasian dengan metode probabilitas dan statistik untuk memprediksi peluang di masa yang akan datang berdasarkan pengalaman dimasa sebelumnya. Pada penelitian ini, pengembangan sistem menggunakan metode CRISP-DM agar pengerjaan menjadi lebih terurut dan pengujian dilakukan dengan metode blackbox untuk menguji setiap fungsi yang ada pada aplikasi yang dibangun. Dari penelitian ini, menghasilkan nilai probabilitas sebesar 0.001042 dengan tingkat akurasi sebesar 99%. Hasil tersebut, membuktikan bahwa metode Naive Bayes dapat digunakan untuk membantu dalam prediksi tingkat kelulusan peserta sertifikasi (word dan excel), karena menghasilkan tingkat keakuratan yang cukup tinggi. sehingga peserta dapat menentukan program sertifikasi yang tepat dipilih sebelum melakukan uji sertifikasi.

Kata kunci: Prediksi, Sertifikasi, Naïve Bayes, CRISP-DM, BlackBox

1. PENDAHULUAN

1.1. Latar Belakang

Teknologi saat ini sangat pesat perkembangannya, hampir semua dari bidang memakai teknologi yang berbasis komputer. Untuk itu, sebagai orang-orang yang bekerja diruang lingkup teknologi informasi haruslah meningkatkan profesionalisme agar profesi tidak diambil dan dikuasai oleh orang-orang yang bukan dari dasarnya orang teknologi informasi. Teknologi informasi selalu berkembang baik secara revolusioner seperti misalnya perkembangan dunia perangkat keras maupun yang lebih bersifat evolusioner seperti yang terjadi pada perkembangan perangkat lunak.

Sertifikasi dalam dunia teknologi informasi merupakan hal yang sangat penting. Karena sertifikasi di bidang teknologi informasi memberikan kredibilitas bagi pemegangnya. Sertifikasi teknologi informasi menunjukkan para Professional teknologi informasi memiliki pengetahuan dan kompetensi yang dapat dibuktikan. Sertifikasi teknologi informasi juga memberikan keunggulan bersaing bagi perusahaan, khususnya dalam pasar global karena kemampuan dan pengetahuan profesional teknologi informasi telah diuji dan didokumentasikan. Seperti pada umumnya, yang lebih penting dari sertifikasi itu sendiri adalah kemampuan yang sebenarnya dan pengetahuan yang mendasar diberbagai informasi mengenai bidang tersebut. Banyak alasan untuk mendapatkan sertifikasi IT (Information Technology).

Laboratorium ITCC (*Information Technology Certification Center*) merupakan satu diantara beberapa laboratorium komputer yang ada di kampus Sekolah Tinggi Teknik – PLN, yang bergerak dalam bidang sertifikasi dan telah bekerjasama dengan vendor skala internasional seperti Microsoft. Laboratorium ITCC (*Information Technology Certification Center*) telah memiliki memiliki hak otorisasi (Authorized Testing Center) dari pihak certiport untuk penyelenggaraan pelatihan dan sertifikasi (tes) Internasional dari program aplikasi, *Microsoft Office Specialist (MOS)*, *Microsoft Technology Associate (MTA)*, *Microsoft Certified Educator (MCE)*, *Adobe Certified Associate (ACA)*, *Autodesk Certified User (ACU)*.

Laboratorium ITCC (*Information Technology Certification Center*) berdiri pada tahun 2016 hingga 2018, dan telah melakukan uji sertifikasi terhadap peserta dengan program sertifikasi microsoft office word 2013 dan microsoft office excel 2013 dari tahun 2017 dan terus berlanjut. Karena tergolong sebuah laboratorium yang baru, untuk saat ini para peserta sertifikasi dari pihak adalah peserta internal (peserta dari dalam ruang lingkup kampus STT – PLN) sehingga untuk perencanaan kedepannya Laboratorium ITCC (*Information Technology Certification Center*) melakukan sertifikasi dari pihak eksternal (peserta dari luar ruang lingkup kampus STT - PLN).

Terdapat kekurangan dalam kegiatan uji sertifikasi, kekurangan dari kegiatan tersebut adalah masalah memprediksi penentuan kelulusan uji sertifikasi. Sistem tidak ada yang mengetahui masalah prediksi tingkat kelulusan peserta baik dari pihak penyelenggara maupun peserta. Kurangnya pengetahuan peserta terhadap kriteria yang dimiliki untuk memilih program sertifikasi yang tepat antara Microsoft office word 2013 dan Microsoft office excel 2013 serta pemahaman kriteria peserta sebagai penunjang untuk mengetahui tingkat kelulusan, sehingga perlu adanya informasi mengenai prediksi tingkat kelulusan guna memperbaiki peluang untuk lulus pada program sertifikasi yang tepat.

Dikarenakan adanya kekurangan, maka penelitian ini dilakukan untuk menangani kekurangan tersebut. Untuk memudahkan peserta dalam menentukan tingkat kelulusan sertifikasi, maka dilakukan teknik data mining menggunakan metode klasifikasi dengan menerapkan algoritma naïve bayes didalamnya, algoritma naïve bayes merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya sehingga dikenal sebagai teorema Bayes.

Teorema tersebut dikombinasikan dengan "naive" dimana diasumsikan kondisi antar atribut saling bebas.

Naive Bayes ini mampu memprediksi peluang tingkat kelulusan di masa depan sehingga mempermudah dalam peserta memilih program sertifikasi yang tepat. Sistem ini dapat membantu peserta untuk mendapatkan informasi mengenai tingkat kelulusan peserta berdasarkan kriteria yang dimiliki dan memudahkan peserta untuk menentukan program yang dipilih sehingga efektif dan efisien. Penelitian ini nantinya diharapkan akan meminimalisir tingkat ketidaklulusan peserta dalam hal kegiatan uji sertifikasi dan mampu memberikan informasi kepada peserta untuk memilih program sertifikasi yang tepat. Berdasarkan pada latar belakang tersebut, maka penelitian ini dilakukan sebagai prediksi tingkat kelulusan peserta program sertifikasi word dan excel dengan menggunakan algoritma naive bayes pada laboratorium ITCC .

2. LANDASAN TEORI

2.1. Data Mining

Data Mining merupakan suatu proses menemukan hubungan yang berarti kecenderungan dengan memeriksa sekumpulan besar data yang tersimpan dalam penyimpanan dengan teknik pengenalan pola seperti teknik statistik dan matematika. Ketersediaan data yang melimpah, kebutuhan akan informasi atau pengetahuan sebagai pendukung pengambilan keputusan untuk membuat solusi bisnis, dan dukungan infrastruktur di bidang teknologi informasi merupakan pendukung dari pemrosesan data mining ini. Data Mining dimanfaatkan untuk mendeteksi kejadian-kejadian yang ganjil seperti berbagai penerapan data mining yakni Analisa pasar dan manajemen, Analisa perusahaan dan manajemen resiko, Telekomunikasi, Keuangan, Asuransi, Olahraga dan Astronomi. Sebagai suatu teknologi yang dapat menghasilkan knowledge, data mining melakukan pemrosesan dalam beberapa tahap [1].

2.2. Naïve Bayes

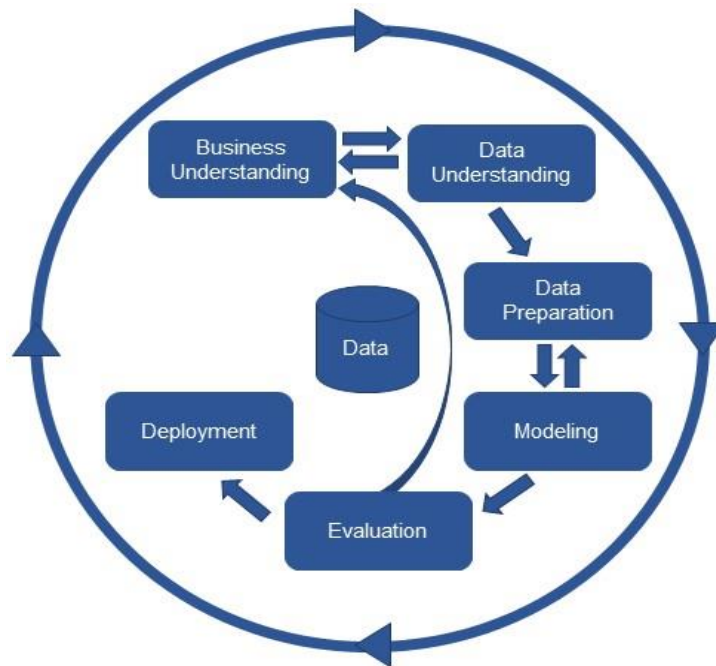
Naive Bayes merupakan pengklasifikasian probabilistik sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan. Algoritma menggunakan teorema Bayes dan mengasumsikan semua atribut independen atau tidak saling ketergantungan yang diberikan oleh nilai pada variabel kelas [2]. Definisi lain mengatakan Naive Bayes merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya. Naive Bayes didasarkan pada asumsi penyederhanaan bahwa nilai atribut secara kondisional saling bebas jika diberikan nilai output. Dengan kata lain, diberikan nilai output, probabilitas mengamati secara bersama adalah produk dari probabilitas individu. Keuntungan penggunaan Naive Bayes adalah bahwa metode ini hanya membutuhkan jumlah data pelatihan (Training Data) yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian. Naive Bayes sering bekerja jauh lebih baik dalam kebanyakan situasi dunia nyata yang kompleks dari pada yang diharapkan.

2.3. CRIPS-DM

Cross-Industry Standart Process for Data Mining (CRISP-DM) yang di kembangkan tahun 1996 oleh analisis dari beberapa industri seperti Daimler Chrysler, SPSS dan NCR. CRISP-DM menyediakan standar proses data mining sebagai strategi pemecahan masalah secara umum dari bisnis atau unit penelitian.

Dalam CRISP-DM sebuah proyek data mining memiliki siklus hidup yang terbagi dalam enam fase Gambar Keseluruhan fase berurutan yang ada tersebut bersifat adaptif. Fase berikutnya dalam urutan bergantung kepada keluaran dari fase sebelumnya. Hubungan penting antar fase digambarkan dengan panah. Sebagai contoh, jika proses berada pada fase modeling. Berdasar pada perilaku dan karakteristik model, proses mungkin kembali kepada fase data preparation untuk perbaikan lebih lanjut terhadap data atau berpindah maju kepada fase evaluation [3].

Terdapat enam fase dalam CRISP-DM ini yakni dijelaskan pada gambar berikut:



Gambar 1. Fase CRISP-DM

3. METODE PENELITIAN

3.1. Analisa Kebutuhan

Analisa kebutuhan merupakan tahapan awal penelitian. Pada tahap ini dilakukan analisa kebutuhan suatu data menggunakan metode CRISP-DM, dimana data akan diproses dengan melakukan beberapa tahapan seperti berikut:

a. *Business Understanding Phase* (Fase Pemahaman Bisnis)

Pada tahap ini dilakukan penentuan tujuan proyek dan kebutuhan secara detail dalam lingkup bisnis atau unit penelitian secara keseluruhan. Penelitian ini memiliki tujuan untuk menciptakan sistem penentuan yang dapat memprediksi tingkat kelulusan peserta sertifikasi berdasarkan kriteria uji sertifikasi dengan mengimplementasikan pendekatan Algoritma Naive Bayes. Untuk mencapai tujuan tersebut dilakukan analisa terlebih dahulu.

1. Analisa Masalah

Analisa permasalahan ini dilakukan untuk mengetahui masalah-masalah yang terjadi dengan menggunakan metode Fishbone. Fishbone atau tulang ikan menjelaskan sebuah sebab-akibat dari

sebuah masalah. Berikut merupakan fishbone yang telah dibuat oleh peneliti, yaitu sebagai berikut:



Gambar 2. Diagram Fishbone

Diagram fishbone menjelaskan tentang sebab – akibat dalam suatu masalah. Adapun penjelasan dari masing-masing penyebabnya adalah:

- Manusia: Penyebab sulitnya menentukan program sertifikasi yang tepat untuk dipilih, yaitu karena kurangnya pengetahuan para peserta mengenai kriteria uji sertifikasi yang dimiliki dan tidak adanya ilmu tentang prediksi. Sehingga peserta sering kali tidak tepat dalam menentukan program yang tepat.
- Metode: Dari pihak laboratorium masih kurang dalam memberikan edukasi mengenai program sertifikasi yang tepat untuk dipilih. Hal ini disebabkan karena belum adanya metode untuk mendukung proses prediksi.
- Sistem: Belum ada sistem untuk memprediksi tingkat kelulusan uji sertifikasi sehingga kurangnya informasi terkait uji sertifikasi
- Material: Hasil data uji sertifikasi sebelumnya tidak diolah sehingga data yang tersimpan dibiarkan begitu
- saja dan peserta tidak mengetahui kriteria yang dimiliki.

Berdasarkan analisa diatas, penulis akan membuat sebuah sistem aplikasi yang akan membantu dalam pengolahan data yang lebih efektif dan efisien, agar data tersebut dapat diolah dan menghasilkan sistem yang bisa membantu proses prediksi dalam menentukan tingkat kelulusan peserta sertifikasi.

b. *Data Understanding Phase (Fase Pemahaman Data)*

Pada tahap ini akan dilakukan pengumpulan data. Pada penelitian kali ini, data yang digunakan adalah data hasil uji sertifikasi di laboratorium ITCC pada tahun 2017 – 2018 dengan jumlah sebanyak 1024 data. Data yang diperoleh adalah sebagai berikut:

Tabel 1. Data Hasil Uji Sertifikasi

No	Program	Kriteria 1	Kriteria 2	Kriteria 3	Kriteria 4	Kriteria 5	Keterangan	Nilai
1	MOS-01	100%	100%	100%	100%	100%	Lulus	1000
2	MOS-01	93%	94%	98%	92%	85%	Lulus	924
3	MOS-01	96%	93%	97%	96%	68%	Lulus	900
4	MOS-01	93%	97%	99%	95%	61%	Lulus	890
5	MOS-01	93%	90%	94%	83%	91%	Lulus	902
..
..
..
..
1022	MOS-01	58%	58%	57%	61%	84%	Tidak	637
1023	MOS-01	57%	63%	55%	61%	71%	Tidak	614
1024	MOS-01	56%	58%	61%	64%	61%	Tidak	600

c. *Data Preparation Phase* (Fase Pengolahan Data)

Pada tahap pengolahan data dilakukan proses data selection dan data transformation. Data yang sudah digabungkan akan diolah di fase pengolahan data ini.

1. *Data Selection* (Seleksi Data)

Proses penyeleksian data dilakukan karena tidak semua data akan digunakan, sehingga dibutuhkan seleksi data. Data yang diambil adalah data yang sesuai dengan analisis yang digunakan. Peneliti hanya menggunakan beberapa data, yaitu sebagai berikut:

Tabel 2. *Data Selection*

No	Kriteria 1	Kriteria 2	Kriteria 3	Kriteria 4	Kriteria 5	Keterangan	Nilai
1	100%	100%	100%	100%	100%	Lulus	1000
2	93%	94%	98%	92%	85%	Lulus	924
3	96%	93%	97%	96%	68%	Lulus	900
4	93%	97%	99%	95%	61%	Lulus	890
5	93%	90%	94%	83%	91%	Lulus	902
..
..
..
..
1022	58%	58%	57%	61%	84%	Tidak	637
1023	57%	63%	55%	61%	71%	Tidak	614
1024	56%	58%	61%	64%	61%	Tidak	600

2. *Data Transformation* (Pengubahan Data)

Data yang telah di seleksi selanjutnya akan di proses lagi pada tahap pengubahan data. Tahap pengubahan data merupakan penggabungan data dengan melakukan pengklasifikasian menjadi beberapa kategori. Berikut tabel pengklasifikasian data:

a) Kriteria 1

Variabel ini berisi range nilai yang telah di klasifikasikan menjadi 5 Kategori, yaitu A, B, C, D dan E. Berikut kriteria 1 yang telah di klasifikasi sebagai berikut:

Tabel 3. Klasifikasi Kriteria

Kriteria1	Kategori
90%-100%	A
80%-89%	B
65%-79%	C
55%-64%	D
<55%	E

b) Kriteria2

Variabel ini berisi range nilai yang telah di klasifikasikan menjadi 5 Kategori, yaitu A, B, C, D dan E. Berikut kriteria2 yang telah di klasifikasi sebagai berikut:

Tabel 4. Klasifikasi Kriteria 2

Kriteria2	Kategori
90%-100%	A
80%-89%	B
65%-79%	C
55%-64%	D
<55%	E

c) Kriteria3

Variabel ini berisi range nilai yang telah di klasifikasikan menjadi 5 Kategori, yaitu A, B, C, D dan E. Berikut kriteria3 yang telah di klasifikasi sebagai berikut:

Tabel 5. Klasifikasi Kriteria 3

Kriteria3	Kategori
90%-100%	A
80%-89%	B
65%-79%	C
55%-64%	D
<55%	E

d) Kriteria4

Variabel ini berisi range nilai yang telah di klasifikasikan menjadi 5 Kategori, yaitu A, B, C, D dan E. Berikut kriteria4 yang telah di klasifikasi sebagai berikut:

Tabel 6. Klasifikasi Kriteria 4

Kriteria4	Kategori
90%-100%	A
80%-89%	B
65%-79%	C
55%-64%	D
<55%	E

e) Kriteria5

Variabel ini berisi range nilai yang telah di klasifikasikan menjadi 5 Kategori, yaitu A, B, C, D dan E. Berikut kriteria5 yang telah di klasifikasi sebagai berikut:

Tabel 7. Klasifikasi Kriteria 5

Kriteria5	Kategori
90%-100%	A
80%-89%	B
65%-79%	C
55%-64%	D
<55%	E

f) **Tingkat Kelulusan**

Variabel tingkat keberhasilan merupakan class atau data yang berfungsi untuk menentukan hasil keputusan. Dalam pengelompokan data sudah ditentukan secara tetap agar tidak terjadi kesalahan dalam proses perhitungan program. Data keputusan tingkat keberhasilan memiliki dua nilai yaitu “LULUS” dan “TIDAK”.

Berikut ini merupakan data yang telah di klasifikasi berdasarkan kategori masing-masing atribut:

Tabel 8. Tabel Data *Transformation*

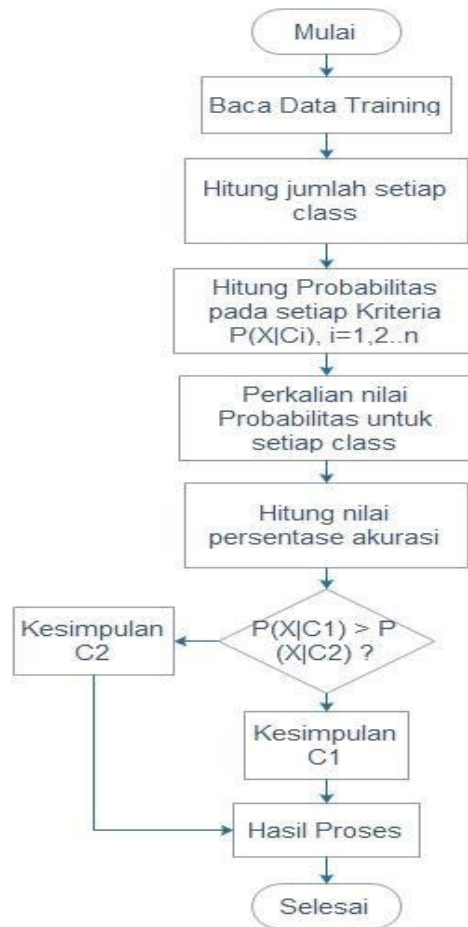
No	Kriteria 1	Kriteria 2	Kriteria 3	Kriteria 4	Kriteria 5	Keterangan
1	A	A	A	A	A	Lulus
2	A	A	A	A	B	Lulus
3	A	A	A	A	C	Lulus
4	A	A	A	A	D	Lulus
5	A	A	A	B	A	Lulus
..
..
..
..
1022	D	D	D	D	D	Tidak
1023	D	D	D	D	D	Tidak
1024	D	D	D	D	D	Tidak

d. *Modelling Phase (Fase Permodelan)*

Pada tahap ini data yang telah di klasifikasikan akan di proses ke tahap selanjutnya dengan beberapa permodelan. Permodelan ini dilakukan untuk mengoptimalkan hasil yang ingin dicapai.

1. *Algoritma Naïve Bayes*

Setelah melakukan semua tahap pemrosesan data, maka akan dihasilkan data training. Data training merupakan data yang akan digunakan untuk melakukan proses perhitungan menggunakan metode Naive Bayes. Berikut gambaran proses perhitungan menggunakan metode Naive Bayes dalam sebuah flowchat, yaitu:



Gambar 3. Flowchart Naïve Bayes

Dari tahapan tersebut, dilakukan perhitungan Algoritma Naive Bayes menggunakan data training dan data testing sebagai data uji. Berikut perhitungan menggunakan Algoritma Naive Bayes:

- a) Baca Data Training

Berikut merupakan data training yang digunakan untuk melakukan proses perhitungan.

Tabel 9. Tabel *Training*

No	Kriteria 1	Kriteria 2	Kriteria 3	Kriteria 4	Kriteria 5	Keterangan
1	A	A	A	A	A	Lulus
2	A	A	A	A	B	Lulus
3	A	A	A	A	C	Lulus
4	A	A	A	A	D	Lulus
5	A	A	A	B	A	Lulus
..
..
..
..
1022	D	D	D	D	D	Tidak
1023	D	D	D	D	D	Tidak
1024	D	D	D	D	D	Tidak

Dari data training diatas, akan di uji menggunakan data testing sebagai berikut, yaitu:

Tabel 10. Tabel *Testing*

Nama	Kriteria1	Kriteria2	Kriteria3	Kriteria4	Kriteria5	Tingkat Kelulusan
Bowval Valentino	90	88	78	85	75	?

Data testing diatas akan di klasifikasi berdasarkan data preparation (data transformation), yaitu:

Tabel 11. Tabel *Testing Transformation*

Nama	Kriteria1	Kriteria2	Kriteria3	Kriteria4	Kriteria5	Tingkat Kelulusan
Bowval Valentino	A	B	C	B	C	?

b) Hitung $P(C1)$ pada setiap Class

Untuk menghitung Probabilitas class, yaitu:

$$P(C1) = \frac{\text{Classtingkatkelulusan} = \text{LULUS}}{\text{Jumlahdatauji}}$$

$$P(C2) = \frac{\text{Classtingkatkelulusan} = \text{TIDAK}}{\text{Jumlahdatauji}}$$

- 1) $P(\text{LULUS}) = \frac{916}{1024} = 0.894531$
- 2) $P(\text{TIDAK}) = \frac{108}{1024} = 0.105469$

c) Hitung probabilitas pada setiap kriteria

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

- 1) Kriteria1 = A
 - i. $P(\text{Kriteria1} = \text{A} | \text{Tingkat kelulusan} = \text{YA}) = \frac{249}{916} = 0.271834$
 - ii. $P(\text{Kriteria1} = \text{A} | \text{Tingkat kelulusan} = \text{TIDAK}) = \frac{7}{108} = 0.064815$
- 2) Kriteria2 = B
 - i. $P(\text{Kriteria2} = \text{B} | \text{Tingkat kelulusan} = \text{YA}) = \frac{243}{916} = 0.265284$
 - ii. $P(\text{Kriteria2} = \text{B} | \text{Tingkat kelulusan} = \text{TIDAK}) = \frac{13}{108} = 0.120370$
- 3) Kriteria3 = C
 - i. $P(\text{Kriteria3} = \text{C} | \text{Tingkat kelulusan} = \text{YA}) = \frac{227}{916} = 0.247817$
 - ii. $P(\text{Kriteria3} = \text{C} | \text{Tingkat kelulusan} = \text{TIDAK}) = \frac{29}{108} = 0.268519$
- 4) Kriteria4 = B
 - i. $P(\text{Kriteria4} = \text{B} | \text{Tingkat kelulusan} = \text{YA}) = \frac{241}{916} = 0.265284$
 - ii. $P(\text{Kriteria4} = \text{B} | \text{Tingkat kelulusan} = \text{TIDAK}) = \frac{13}{108} = 0.120370$
- 5) Kriteria5 = C
 - i. $P(\text{Kriteria5} = \text{C} | \text{Tingkat kelulusan} = \text{YA}) = \frac{222}{916} = 0.245633$
 - ii. $P(\text{Kriteria5} = \text{C} | \text{Tingkat kelulusan} = \text{TIDAK}) = \frac{56}{108} = 0.287037$

d) Perkalian nilai probabilitas pada setiap class

1) Probabilitas LULUS

$$\frac{\frac{\frac{\frac{\frac{916}{1024} * 249}{916} * 243}{916} * 227}{916} * 241}{916} * 222$$

$$0.894531 * 0.271834 * 0.265284 * 0.247817 * 0.265284 * 0.245633$$

$$0.001042$$

2) Probabilitas TIDAK

$$\frac{\frac{\frac{\frac{\frac{108}{1024} * 7}{108} * 13}{108} * 29}{108} * 13}{108} * 56$$

$$0.105469 * 0.064815 * 0.120370 * 0.268519 * 0.120370 * 0.287037$$

$$0.000008$$

e) Hitung Persentase Keakurasian Kelulusan

1) Akurasi Persentase LULUS

$$\left(\frac{nLULUS}{nTIDAK + nLULUS} \right) * 100\%$$

$$\left(\frac{0.001042}{0.000008 + 0.001042} \right) * 100\%$$

$$99.24\%$$

2) Akurasi Persentase TIDAK

$$\left(\frac{nTIDAK}{nLULUS + nTIDAK} \right) * 100\%$$

$$\left(\frac{0.000008}{0.001042 + 0.000008} \right) * 100\%$$

$$0.76\%$$

f) Membandingkan Nilai Akurasi

Dari hasil tersebut menyatakan $P(X|C1) > P(X|C2)$ atau $P(X|C2) < P(X|C1)$

g) Hasil

Berdasarkan data training yang berhasil diklasifikasikan 1024 data training yang di uji pada laboratorium. Metode Naive Bayes berhasil memprediksi tingkat kelulusan peserta sertifikasi dengan persentase keakuratan sebesar 99.24%.

e. *Evaluation Phase* (Fase Evaluasi)

Fase evaluasi merupakan tahap untuk mengetahui apakah model yang di rancang telah sesuai atau belum dengan tujuan pada fase awal. Tujuan awal di dirancangnya model ini yaitu agar menghasilkan nilai akurasi yang tinggi, sehingga dapat membuktikan bahwa penelitian yang dilakukan telah berhasil. Jika hasil tersebut belum sesuai, maka akan dilakukan analisa lebih lanjut untuk menghasilkan rekomendasi perbaikan atau perubahan dalam menentukan hasil tingkat kelulusan peserta sertifikasi. Dengan demikian hal tersebut dapat membantu pihak laboratorium dan para calon peserta sertifikasi untuk menentukan tingkat kelulusan peserta sertifikasi serta menentukan program sertifikasi yang tepat terhadap sistem prediksi yang sudah tersedia.

f. *Deployment Phase* (Fase Penyebaran)

Tahapan penyebaran merupakan tahap implementasi untuk pembangunan aplikasi berupa gambaran pengetahuan yang telah diperoleh sehingga nantinya sistem ini dapat digunakan oleh

pengguna lain. dalam hal ini tahapan penyebaran data mining dapat dilakukan pada seluruh calon peserta uji sertifikasi di Laboratorium ITCC. Pada penelitian ini tahapan deployment atau tahap penyebaran tidak dilakukan.

4. HASIL DAN PEMBAHASAN

4.1. Hasil Rancangan Aplikasi

Aplikasi Tingkat Kelulusan Peserta Program Sertifikasi *Word* dan *Excel* dapat digunakan oleh dua *user* yaitu peserta sebagai user biasa dan pihak pengelola laboratorium ITCC sebagai admin. Setiap *user* memiliki hak akses tersendiri terhadap aplikasi. Algoritma *Naïve Bayes* bertujuan untuk mencari tingkat kelulusan peserta dengan membandingkan dua program sertifikasi antara *word* dan *excel*.

#	Kriteria1	Kriteria2	Kriteria3	Kriteria4	Kriteria5	KETERANGAN	AKSI
1	100	100	100	100	100	LULUS	Edit Hapus
2	93	94	98	92	85	LULUS	Edit Hapus
3	96	93	97	96	68	LULUS	Edit Hapus

Gambar 4. Hasil Aplikasi Training

#	NAMA	G1	G2	G3	G4	G5	KETERANGAN	TINGKAT LULUS	TINGKAT TIDAK	LULUS %	AKSI
1	admin	A	B	A	B	A	LULUS	0.001265	0.000000	100.00	Hapus
2	bowwal	A	B	C	B	C	LULUS	0.001042	0.000008	99.24	Hapus

Gambar 5. Hasil Aplikasi Testing

4.2. Pembahasan

Aplikasi prediksi tingkat kelulusan peserta sertifikasi yang dibangun ini berfungsi untuk membantu dalam penentuan program sertifikasi terhadap para calon peserta. Prediksi dilakukan

menggunakan data training sebanyak 1024 data training yang telah diproses sebelumnya dengan beberapa tahapan seperti proses data selection, dan proses data transformation. Data training ini akan digunakan untuk proses perhitungan menggunakan metode Naive Bayes pada data testing yang akan diuji. Pengujian ini bertujuan untuk melihat berapa besar tingkat keakuratan yang akan dihasilkan dari prediksi yang telah dilakukan.

Sebelumnya belum ada aplikasi yang membantu dalam prediksi penentuan tingkat kecocokan jenis tanaman sayuran, karena data yang ada masih diolah secara manual, sehingga sulit untuk diterapkan suatu metode dalam pengolahan data.

Berdasarkan penelitian yang dilakukan menggunakan pendekatan Naive Bayes, nilai probabilitas class LULUS yang dihasilkan sebesar 0.89453125 dan class TIDAK sebesar 0.10546875 dengan tingkat akurasi sebesar 99%. Dari hasil tersebut menunjukkan bahwa peserta sertifikasi dengan kriteria1, kriteria2, kriteria3, kriteria4 dan kriteria5 sebagai data testing memiliki tingkat akurasi yang lulus uji sertifikasi lebih besar dibandingkan dengan akurasi yang tidak lulus uji sertifikasi.

KESIMPULAN

Berdasarkan hasil penelitian dan pembahasan yang telah dilakukan dapat diambil beberapa kesimpulan, antara lain:

1. Aplikasi prediksi menggunakan algoritma naïve bayes dapat memberikan informasi terkait prediksi peserta sertifikasi yang dinyatakan lulus dan tidak lulus sehingga peserta dan pihak pengelola laboratorium ITCC dapat mengetahui seberapa besar tingkat kelulusan sertifikasi.
2. Aplikasi dapat memberikan solusi keputusan untuk memilih program sertifikasi yang tepat dengan membandingkan hasil prediksi uji sertifikasi antara word dan excel.

DAFTAR PUSTAKA

- [1] Larose D, T., 2005, *Discovering knowledge in data: an introduction to data mining*, Jhon Wiley & Sons Inc.
- [2] Syarli, & Muin, A. A. (2016). Metode Naive Bayes Untuk Prediksi Kelulusan (Studi Kasus: Data Mahasiswa Baru Perguruan Tinggi). *Jurnal Ilmiah Ilmu Komputer*, 2(1), 1–5.
- [3] Budiman, I., Prahasto, T., & Christyono, Y. (2012). Data Clustering Menggunakan Metodologi Crisp-Dm Untuk Pengenalan Pola Proporsi Pelaksanaan Tridharma, 2012(Snati), 15–16.
- [4] A.S, Rosa dan Shalauhuddin, M. (2014). *Rekayasa Perangkat Lunak Terstruktur dan Berorientasi Objek*. Bandung: Informatika Bandung.
- [5] Purwanto, S. D., & Santoso, I. B. (2017). Sistem Identifikasi Boraks pada Bakso Daging Sapi Berbasis Android Menggunakan Algoritma Naive Bayes Classifier, 9(1), 33–37.
- [6] Nurrohmat, M. A., & Nugroho, Y. S. (2015). Aplikasi Pemrediksi Masa Studi dan Predikat Kelulusan Mahasiswa Informatika Universitas Muhammadiyah Surakarta Menggunakan Metode Naive Bayes. *Khazanah Informatika*, I(1), 29–34. <https://doi.org/10.23917/khif.v1i1.1179>
- [7] Siregar, R. R. A., Sinaga, F. A., & Arianto, R. (2017). Aplikasi Penentuan Dosen Penguji Skripsi Menggunakan Metode TF-IDF dan Vector Space Model. *Computatio: Journal of Computer Science and Information Systems*, 1(2), 171-186.
- [8] T. Nisa, R. Siregar, And W. Suliyanti, “Estimasi Daya Beban Listrik Pada Gardu Induk Cengkareng Dengan Menggunakan Metode Time Series Model Dekomposisi”, *Teknologia*, Vol. 1, No. 2, Apr. 2019.

- [9] Sangadji, I., & Arvio, Y. (2018, March). Dynamic Segmentation Of Behavior Patterns Based On Quantity Value Movement Using Fuzzy Subtractive Clustering Method. In *Journal of Physics: Conference Series* (Vol. 974, No. 1, p. 012009). IOP Publishing.
- [10] Siregar, R., Siregar, Z., & Arianto, R. (2019). Klasifikasi Sentiment Analysis Pada Komentar Peserta Diklat Menggunakan Metode K-Nearest Neighbor. *KILAT*, 8(1). <https://doi.org/10.33322/kilat.v8i1.421>
- [11] A. Prianty, R. Siregar, and R. Arianto, “Penanganan Gangguan Listrik Rumah Tangga Menggunakan Algoritma Greedy Untuk Penentuan Jarak Optimal”, *Teknologi*, vol. 2, no. 1, Aug. 2019.
- [12] Siregar Rr, Putri Dr. Metode Support Vector Machine Pada Klasifikasi Audit Energi: Studi Kasus Gedung STT-PLN Jakarta. *Jurnal Informatika Dan Komputasi*. 2017 Mar 2;8(2):98-104.
- [13] R. Jurnal, “Implementasi Logika Fuzzy Untuk Sistem Otomatisasi Pengaturan Pengisian Baterai Pembangkit Listrik Tenaga Surya”, *Energi*, Vol. 9, No. 2, Pp. 111-119, Nov. 2018.