

**KLASIFIKASI LINEARITAS SERAPAN LULUSAN SEKOLAH  
MENENGAH KEJURUAN MENGGUNAKAN  
ALGORITMA *NAIVE BAYES***

**THESIS**

**Oleh:  
HIRGA ERTAMA PUTRA  
NIM. 200605210007**



**PROGRAM STUDI MAGISTER INFORMATIKA  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM  
MALANG  
2024**

**KLASIFIKASI LINEARITAS SERAPAN LULUSAN SEKOLAH  
MENENGAH KEJURUAN MENGGUNAKAN  
ALGORITMA *NAIVE BAYES***

**THESIS**

**Diajukan kepada:**

**Fakultas Sains dan Teknologi  
Universitas Islam Negeri Maulana Malik Ibrahim Malang  
Untuk memenuhi Salah Satu Persyaratan dalam  
Memperoleh Gelar Magister Komputer (M.Kom)**

**Oleh:**

**HIRGA ERTAMA PUTRA  
NIM. 200605210007**

**PROGRAM STUDI MAGISTER INFORMATIKA  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM  
MALANG  
2024**

**KLASIFIKASI LINEARITAS SERAPAN LULUSAN SEKOLAH  
MENENGAH KEJURUAN MENGGUNAKAN  
ALGORITMA *NAIVE BAYES***

**TESIS**

Diajukan kepada:

Fakultas Sains dan Teknologi  
Universitas Islam Negeri Maulana Malik Ibrahim Malang  
Untuk memenuhi Salah Satu Persyaratan dalam  
Memperoleh Gelar Magister Komputer (M.Kom)

Oleh:

**HIRGA ERTAMA PUTRA  
NIM. 200605210007**

**PROGRAM STUDI MAGISTER INFORMATIKA  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM  
MALANG  
2024**

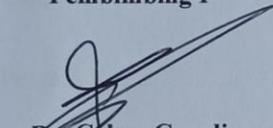
**KLASIFIKASI LINEARITAS SERAPAN LULUSAN SEKOLAH  
MENENGAH KEJURUAN MENGGUNAKAN  
ALGORITMA *NAIVE BAYES***

**THESIS**

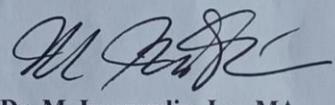
Oleh:  
**HIRGA ERTAMA PUTRA**  
**NIM. 200605210007**

Tesis diperiksa dan disetujui untuk diuji :  
Tanggal 3 Juni 2024

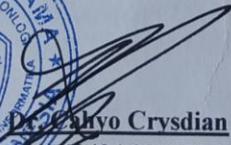
**Pembimbing I**

  
**Dr. Cahyo Crvsdian**  
**NIP. 19740424 200901 1 008**

**Pembimbing II**

  
**Dr. M. Imamudin, Lc., MA.**  
**NIP. 19740602 200901 1 010**

Mengetahui,  
Ketua Program Studi Magister Informatika  
Fakultas Sains dan Teknologi  
Universitas Islam Negeri Maulana Malik Ibrahim Malang

  
  
**Dr. Cahyo Crvsdian**  
**NIP. 19740424 200901 1 008**

**KLASIFIKASI LINEARITAS SERAPAN LULUSAN SEKOLAH  
MENENGAH KEJURUAN MENGGUNAKAN  
ALGORITMA *NAIVE BAYES***

**THESIS**

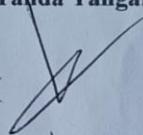
**Oleh:  
HIRGA ERTAMA PUTRA  
NIM. 200605210007**

Telah Dipertahankan di Depan Dewan Penguji Thesis  
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan  
Untuk Memperoleh Gelar Magister Komputer (M.Kom)  
Tanggal 3 Juni 2024

**Susunan Dewan Penguji**

**Tanda Tangan**

Penguji I : Dr. Irwan Budi Santoso, M.Kom.  
NIP. 19770103 201101 1 004

(  )

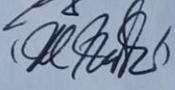
Penguji II : Dr. Muhammad Faisal, M.T.  
NIP 19740510 200501 1 007

(  )

Pembimbing I : Dr. Cahyo Crysdian  
NIP 19740424 200901 1 008

(  )

Pembimbing II : Dr. M. Imamudin, Lc., MA.  
NIP. 19740602 200901 1 010

(  )

Mengetahui dan Mengesahkan  
Ketua Program Studi Magister Informatika  
Fakultas Sains dan Teknologi  
Universitas Islam Negeri Maulana Malik Ibrahim Malang

  
Dr. Cahyo Crysdian  
NIP. 19740424 200901 1 008

## PERNYATAAN KEASLIAN TULISAN

Yang bertanda tangan di bawah ini saya:

Nama : Hirga Ertama Putra  
NIM : 200605210007  
Program : Magister Informatika  
Fakultas : Sains dan Teknologi

Menyatakan dengan sebenarnya bahwa Thesis yang saya tulis ini benar-banar merupakan hasil karya saya sendiri, bukan merupakan pengambilalihan data, tulisan atau pikiran orang lain yang saya akui sebagai hasil tulisan atau pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar pustaka.

Apabila dikemudian hari terbukti atau dapat dibuktikan Thesis ini hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 31 Mei 2024

Saya yang menyatakan,



Hirga Ertama Putra  
NIM. 200605210007

## **MOTTO**

*“Cukuplah bagi kami Allah, sebaik-baiknya pelindung dan sebaik-baiknya penolong kami.”*

*Ali Imran ayat 173 dan Al-Anfal ayat 40*

## PERSEMBAHAN

Dengan mengucapkan syukur *Alhamdulillah rabbil alamin*, Thesis ini saya persembahkan untuk :

1. Allah *Subhanahu Wa Ta'ala* yang memberikan rahmat, hidayah, kesehatan, rejeki, serta semua yang saya butuhkan sampai selesainya tesis ini.
2. Seluruh keluarga tercinta (istriku Oktarica Pratiwi Suryoningtyas, anak pertama Irlande Gata Agistama, anak kedua Hirlanta Hafizh Agistama dan anak ketiga Irpandu Abid Agistama), yang selalu memberikan dukungan, doa dan semangat yang tiada henti.
3. Seluruh orang tua tersayang (bapakku Ir. Bambang Hermanto dan ibukku Ir. Sri Hartati tercinta, serta bapak Drs. Agus Suryo Raharjo, MM.M.Pd dan ibu Siswanti Sulistyو Budi).
4. Dr. Cahyo Crysdiان dan Dr. M. Imammudin, Lc., MA., selaku dosen pembimbing yang sudah banyak memberikan ilmu, arahan dan bimbingannya.
5. Dr. Irwan Budi Santoso, M.Kom. dan Dr. Muhammad Faisal, M.T., selaku tim penguji tesis yang sudah banyak membantu terselesainya tesis saya.
6. Seluruh civitas akademika Universitas Islam Negeri Maulana Malik Ibrahim Malang untuk dapat menimba ilmu, pengalaman dan khususnya tentang ilmu agama.
7. Seluruh civitas akademika dan jajaran tim manajemen SMK Telkom Malang yang memberikan dukungan moril, materiil dan kesempatan untuk dapat studi lanjut.
8. Bapak, ibu, saudara dan rekan-rekan sekalian yang tidak bisa saya sebutkan satu persatu yang telah mendukung penyelesaian tesis ini.

## KATA PENGANTAR

*Assalamu'alaikum Wr. Wb.*

Syukur alhamdulillah penulis hanturkan kehadiran Allah *Subhanahu Wa Ta'ala* yang telah melimpahkan Rahmat dan Hidayah-Nya, sehingga penulis dapat menyelesaikan studi di Program Studi Magister Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang sekaligus menyelesaikan tesis ini dengan baik.

Selanjutnya penulis haturkan ucapan terima kasih seiring do'a dan harapan *jazakumullah ahsanal jaza'* kepada semua pihak yang telah membantu terselesaikannya tesis ini. Ucapan terima kasih ini penulis sampaikan kepada:

1. Bapak Dr. Cahyo Crysdiand dan Bapak Dr. M. Imammudin, Lc., MA., selaku dosen pembimbing yang sudah banyak memberikan ilmu, arahan dan bimbingannya.
2. Segenap civitas akademika Program Studi Magister Informatika, terutama seluruh Bapak/ Ibu dosen, terima kasih atas segenap ilmu dan bimbingannya.
3. Seluruh civitas akademika dan jajaran tim manajemen SMK Telkom Malang, yang sudah banyak membantu terselesaikannya tesis ini.
4. Ayahanda dan Ibunda tercinta yang senantiasa memberikan doa dan restunya kepada penulis dalam menuntut ilmu.
5. Keluarga penulis yang selalu memberikan semangat kepada penulis untuk menyelesaikan tesis ini.
6. Semua pihak yang ikut membantu dalam menyelesaikan tesis ini yang tidak bisa penulis sebutkan satu persatu.

Penulis menyadari bahwa dalam penyusunan tesis ini masih terdapat kekurangan dan penulis berharap semoga tesis ini bias memberikan manfaat kepada para pembaca khususnya bagi penulis secara pribadi. Amin Ya Rabbal Alamin.

*Wassalamu'alaikum Wr. Wb.*

Malang, Mei 2024

Penulis

## DAFTAR ISI

HALAMAN JUDUL .....	i
HALAMAN PENGAJUAN .....	ii
HALAMAN PERSETUJUAN.....	iii
HALAMAN PENGESAHAN .....	iv
HALAMAN PERNYATAAN.....	Error! Bookmark not defined.
MOTTO .....	vii
PERSEMBAHAN.....	viii
KATA PENGANTAR.....	ix
DAFTAR ISI.....	x
DAFTAR TABEL.....	xii
DAFTAR GAMBAR.....	xiii
ABSTRAK .....	xiv
ABSTRACT .....	xv
المخلص.....	xvi
<b>BAB I PENDAHULUAN.....</b>	<b>1</b>
1.1. Latar Belakang Masalah .....	1
1.2. Pernyataan Masalah .....	5
1.3. Tujuan Penelitian.....	5
1.4 Manfaat Penelitian .....	6
1.5 Ruang Lingkup Penelitian.....	6
<b>BAB II KAJIAN PUSTAKA .....</b>	<b>6</b>
2.1 Sistem Klasifikasi Serapan Lulusan .....	7
2.2 Kerangka Teori.....	13
<b>BAB III DESAIN PENELITIAN.....</b>	<b>17</b>
3.1 Prosedur Penelitian .....	17
3.1.1 <i>Data Collection</i> .....	17
3.1.2 <i>System Development</i> .....	21
3.1.3 <i>Feature Extraction</i> .....	22
3.2 Konseptual Model .....	32

<b>BAB IV METODE <i>NAIVE BAYES</i></b> .....	<b>34</b>
<b>4.1 Desain Metode</b> .....	<b>34</b>
<b>4.1.1 Perhitungan Nilai Prior</b> .....	<b>35</b>
<b>4.1.2 Perhitungan Nilai <i>Evidence</i></b> .....	<b>35</b>
<b>4.1.3 Perhitungan Nilai Likelihood</b> .....	<b>39</b>
<b>4.1.4 Perhitungan Nilai Posterior</b> .....	<b>46</b>
<b>4.2 Ujicoba</b> .....	<b>50</b>
<b>4.3 Kesimpulan</b> .....	<b>54</b>
<b>BAB V METODE <i>DECISION TREE</i></b> .....	<b>56</b>
<b>5.1 Desain Metode</b> .....	<b>56</b>
<b>5.2 Ujicoba</b> .....	<b>58</b>
<b>5.3 Kesimpulan</b> .....	<b>68</b>
<b>BAB VI PEMBAHASAN</b> .....	<b>70</b>
<b>6.1 Pembahasan Komparasi Performa Algoritma</b> .....	<b>70</b>
<b>6.2 Klasifikasi linearitas serapan lulusan dalam pandangan agama Islam</b> .....	<b>71</b>
<b>BAB VII KESIMPULAN</b> .....	<b>74</b>
<b>7.1 Kesimpulan</b> .....	<b>74</b>
<b>7.2 Saran</b> .....	<b>75</b>
<b>DAFTAR PUSTAKA</b> .....	<b>76</b>

## DAFTAR TABEL

Tabel 2.1 Kriteria .....	7
Tabel 3.1 Parameter .....	18
Tabel 3.3 Tingkatan Kriteria Parameter .....	22
Tabel 3.4 <i>Dataset</i> Alumni .....	27
Tabel 3.5 <i>Dataset</i> Alumni Cleaning .....	30
Tabel 4.1 <i>Data Testing Naive Bayes</i> .....	51
Tabel 4.2 <i>Confusion Matrix</i> .....	52
Tabel 4.3 Hasil Eksperimen .....	54
Tabel 5.1 Perhitungan Entropy Dan Gain .....	57
Tabel 5.2 <i>Data Testing Decision Tree</i> .....	59
Tabel 5.3 <i>confusion matrix</i> .....	66
Tabel 5.4 Hasil Eksperimen .....	69
Tabel 6.1. Perbandingan performa algoritma <i>Naive Bayes</i> dengan <i>Decision Tree</i> .....	70

## DAFTAR GAMBAR

Gambar 2.1 Kerangka Teori .....	14
Gambar 3.1 Desain Penelitian .....	17
Gambar 3.2 Desain Sistem .....	21
Gambar 3.3 Koseptual Model .....	32
Gambar 4.1 Confusion Matrix linearitas lulusan .....	53

## ABSTRAK

Putra, Hirga Ertama, 2024, **Klasifikasi Linearitas Serapan Lulusan Sekolah Menengah Kejuruan Menggunakan Algoritma *Naive Bayes***, Program Magister Informatika Universitas Islam Negeri Maulana Malik Ibrahim, Pembimbing: (1) Dr. Cahyo Crysdian (2) Dr. H. M. Imamudin, Lc., M.A.

**Kata Kunci :** Klasifikasi, Liniearitas, Serapan Lulusan, *Naive Bayes*, *Decision Tree*.

Pendidikan merupakan instrumen dasar dalam peningkatan kualitas suatu bangsa dan negara. Melalui Sumber Daya Manusia (SDM) yang handal akan menambah kesejahteraan dan peningkatan kualitas bangsa. Secara yuridis SMK dibentuk untuk memberikan alumni yang siap bekerja. Berdasarkan data Badan Pusat Statistik (BPS) menyampaikan bahwa TPT (Tingkat Pengangguran Terbuka) tamatan SMK masih merupakan yang paling tinggi dibandingkan tamatan jenjang pendidikan lainnya. Tercatat pada bulan Februari 2023 sebesar 5,45 persen dari jumlah penduduk Indonesia. Berbeda dengan kondisi yang ada di sekolah, banyak sekolah yang sudah bisa menyalurkan alumninya dengan cara mengadakan kegiatan *jobfair*. Dari hasil *jobfair*, harapannya adalah para alumni dan industri akan saling bertemu untuk melakukan proses *rekrutmen*. Dengan menggunakan *Machine Learning* diharapkan dapat membantu untuk mengklasifikasikan linearitas serapan lulusan siswa SMK. Algoritma *Naive Bayes* dan Algoritma *Decision Tree* dipilih untuk mengolah dataset serapan lulusan SMK sehingga akan menghasilkan performa *accuracy*, *precision*, *recall* dan *f1-score* pada masing - masing algoritma. Dari hasil pengujian algoritma *Naive Bayes* pada strategi eksperimen menghasilkan nilai *accuracy* sebesar 80%, *precision* sebesar 99%, *recall* sebesar 81% dan *f1-score* sebesar 89%, sedangkan pengujian algoritma *Decision Tree* pada strategi eksperimen menghasilkan nilai *accuracy* sebesar 76%, *precision* sebesar 92%, *recall* sebesar 82% dan *f1-score* sebesar 86%. Berdasarkan hasil pengujian eksperimen tersebut disimpulkan bahwa algoritma *Naive Bayes* sebagai metode terbaik karena memiliki nilai *accuracy* lebih baik daripada algoritma *Decision Tree*.

## ABSTRACT

Putra, Hirga Ertama, 2024, **Linearity Classification of Vocational High School Graduates Absorption Using the Naive Bayes Algorithm**, Master Program in Informatics, Maulana Malik Ibrahim State Islamic University, Promotor : (1) Dr. Cahyo Crys dian (2) Dr. H. M. Imamudin, Lc., M.A.

**Keyword** : Classification, Linearity, Graduate Uptake, Naive Bayes, Decision Tree.

Education is a basic instrument in improving the quality of a nation and state. Through reliable Human Resources (HR), it will increase the welfare and improve the quality of the nation. Legally, Vocational Schools were formed to provide alumni who are ready to work. Based on data from the Central Statistics Agency (BPS), the TPT (Open Unemployment Rate) for vocational school graduates is still the highest compared to graduates of other levels of education. It was recorded in February 2023 at 5.45 percent of Indonesia's population. In contrast to the conditions that exist in schools, many schools are able to channel their alumni by holding job fair activities. From the results of the job fair, the hope is that alumni and industry will meet each other to carry out the recruitment process. By using Machine Learning it is hoped that it can help to classify the linearity of absorption of vocational school student graduates. The Naive Bayes Algorithm and Decision Tree Algorithm were chosen to process the uptake dataset of vocational school graduates so that it will produce accuracy, precision, recall and f1-score performance in each algorithm. From the results of testing the Naive Bayes algorithm on the experimental strategy, it produces an accuracy value of 80%, precision of 99%, recall of 81% and f1-score of 89%, while testing the Decision Tree algorithm on the experimental strategy produces an accuracy value of 76%, precision of 92%, recall of 82% and f1-score of 86%. Based on the results of these experimental tests, it was concluded that the Naive Bayes algorithm was the best method because it had better accuracy values than the Decision Tree algorithm

## الملخص

بوترا، هيرجا إرتاما، ٢٠٢٤، التصنيف الخطي لاستيعاب المتخرجي المدرسة المتوسطة المهنية باستخدام خوارزمية نائف بايس ، برنامج ماجستير المعلوماتية، جامعة إسلامية حكومية مولانا مالك إبراهيم ، المشرف: (١) د. جاهيو جريسديان (٢) د. الحج محمد إمام الدين، M.A ،Lc.

الكلمات المفتاحية: التصنيف، الخطية، نائف بايس ، دجسسين تري، استيعاب المخترجين .

التربية هي أداة أساسية لتحسين جودة الأمة والدولة. من خلال الموارد البشرية الموثوقة ستزيد الرفاهية وتحسين جودة الأمة . من الناحية القانونية، تشكل المدارس المتوسطة المهنية لتزويد المتخرجين المستعدين للعمل. لاعتبارة بيانات الجهاز المركزي للإحصاء . (BPS) لا يزال معدل البطالة المفتوحة (TPT) لمتخرجي المدارس المهنية هو الأعلى من المتخرجين من مستويات التربية الأخرى. تسجل في شهر فبراير ٢٠٢٣ سيبلغ ٥,٤٥ ٪ من جمل السكان إندونيسيا. وعلى المختلف من الظروف الموجودة في المدارس، كثير من المدارس قادرة على تصريف متخيجيها من خلال إقامة أنشطة جوفائير. ومن نتائج أنشطة جوفائير، ترجى أن يلتقي الخريجون والصناعة ببعض البعض لتنفيذ عملية ريكروتمين. باستخدام ماجين ليرنيغ ، ترجى أن تساعد في تصنيف خطية استيعاب خريجي المدارس المهنية. اختيار خوارزمية نائف بايس و خوارزمية دجسسسين تري لمعالجة مجموعة بيانات استيعاب خريجي المدارس المتوسطة المهنية حتى تحصل اداء دقة و فرشين و ف١ شكور في كل خوارزميات. من نتائج اختبار خوارزمية نائف بايس على الاستراتيجية التجريبية حصلت على قيمة دقة ٩٢٪ فرشين ٩٢٪ و ريكول بحجم ١٠٠٪ و ف١ شكور ٩٦٪، بينما تم اختبار خوارزمية دجسسسين تري على الاستراتيجية التجريبية تنتج قيمة دقة ٨٩٪، و فرشين ٩٢٪، و ريكول بحجم ٩٧٪، و ف١ شكور ٩٤٪. وبناء على نتائج هذه الاختبارات التجريبية، تم التوصل إلى أن خوارزمية نائف بايس هي الطريقة الأفضل لأنها تتمتع بقيمة دقة أفضل من خوارزمية دجسسسين تري.

# **BAB I**

## **PENDAHULUAN**

### **1.1. Latar Belakang Masalah**

Pendidikan adalah fondasi utama dalam meningkatkan mutu sebuah bangsa dan negara. Dengan Sumber Daya Manusia (SDM) yang kompeten, kesejahteraan dan standar hidup bangsa dapat ditingkatkan. Data yang dirilis oleh Badan Pusat Statistik (BPS) mencatat tingkat pengangguran pada bulan Februari 2023 mencapai 5,45 persen dari total penduduk Indonesia (Said, 2023).

Lulusan Sekolah Menengah Kejuruan (SMK) masih menjadi kelompok penyumbang pengangguran terbanyak, berdasarkan laporan dari CNN Indonesia yang menyatakan bahwa Tingkat Pengangguran Terbuka (TPT) di kalangan lulusan SMK masih mencapai angka tertinggi, yakni sebesar 9,6 persen (CNN Indonesia, 2023). Selain laporan dari CNN, studi sebelumnya juga telah mengungkap indikator yang menyoroti peran SMK sebagai penyumbang tinggi angka pengangguran di Provinsi Jawa Timur (Mukhlason *et al.*, 2020). Penelitian sebelumnya juga menyoroti dinamika permasalahan ketenagakerjaan dan pengangguran di Indonesia (Suhandi *et al.*, 2021). Hasil penelitian tersebut menegaskan bahwa masalah pengangguran merupakan isu yang kompleks, memerlukan kerjasama dari berbagai pihak. Pentingnya aspek kelembagaan sebagai elemen kunci dalam upaya menekan angka pengangguran juga diutarakan dalam penelitian tersebut.

Prestasi dan keberhasilan siswa SMK telah mencapai tingkat yang mengesankan, terbukti dengan sejumlah kejuaraan dalam berbagai kompetisi yang berhasil diraih oleh para siswa SMK. Sebuah laporan dari Kompas menyebutkan

bahwa siswa SMK meraih 15 medali dalam ajang World Skill Competition (WSC) (Hutapea *et al.*, 2019). Selain meraih prestasi dalam kompetisi, siswa SMK juga menunjukkan kemampuan dalam menciptakan inovasi dan karya-karya lainnya. Sebuah berita berjudul "Potret Karya Siswa SMK di Penjuru Negeri" mengulas kreativitas siswa SMK di berbagai wilayah Indonesia (IDN Times, 2023).

Keterserapan siswa SMK juga sangat luar biasa, dibuktikan dalam proses rekrutmen yang dilakukan oleh banyak sekolah. Dengan adanya Job Fair yang diadakan oleh hampir setiap sekolah, seperti SMK PGRI 3 Malang menyebutkan bahwa salah satu tugas kami sebagai SMK PK untuk mengentaskan pengangguran baik dari lulusan Skariga sendiri maupun dari sekolah yang memiliki aliansi keanggotaan FKBKK kota malang (MALANG POSCO MEDIA, 2023). Selain SMK PGRI 3 masih banyak sekolah yang melakukan Job Fair, seperti yang dilakukan oleh SMK Negeri 11 Malang. Diikuti dari 22 industri kreatif dan umum, yang dikabarkan oleh tim media tribun jatim (Widyawati, 2023). SMK Telkom Malang juga melakukan hal yang sama. Tidak hanya melakukan Job Fair tetapi juga memfasilitasi siswa yang berkeinginan wirausaha. Bahkan dimulai dari proses mentoring hingga pameran hasil karya yang dipresentasikan didepan industri dan para investor bisnis (Ramadhan, 2022).

Bursa Kerja Khusus (BKK) adalah unit pelayanan pada SMK, perguruan tinggi, dan Balai Latihan Kerja (BLK) yang memberikan fasilitasi penempatan tenaga kerja kepada semua alumninya (Menteri Ketenagakerjaan Republik Indonesia, 2016). Didalam peraturan tersebut sudah dijelaskan bahwa BKK mempunyai tugas untuk menyalurkan alumninya untuk bekerja. Beberapa kegiatan

dan upaya yang terus dilakukan oleh unit BKK dalam Sekolah Menengah Kejuruan (SMK) yaitu memberikan informasi dan tempat untuk melakukan proses rekrutmen dari industri. Selain itu melakukan penelusuran alumni yang berkerja ataupun yang melanjutkan pendidikan lanjut (Lukman, 2018).

Secara hukum, didirikan Sekolah Menengah Kejuruan (SMK) dengan tujuan menghasilkan lulusan yang telah siap untuk memasuki dunia kerja. Menurut Undang-Undang Negara Republik Indonesia Nomor 20 Tahun 2003 tentang Sistem Pendidikan Nasional, Pasal 18 ayat 2 menyatakan bahwa "Pendidikan menengah terdiri dari pendidikan menengah umum dan pendidikan menengah kejuruan", sementara Pasal 3 menjelaskan bahwa "Pendidikan menengah mencakup Sekolah Menengah Atas (SMA), Madrasah Aliyah (MA), Sekolah Menengah Kejuruan (SMK), dan Madrasah Aliyah Kejuruan (MAK), atau bentuk pendidikan serupa lainnya". Kementerian Pendidikan dan Kebudayaan (Kemendikbud) menjelaskan bahwa SMK bertujuan untuk menciptakan lulusan yang siap bekerja, baik sebagai tenaga kerja yang siap dipekerjakan maupun sebagai pelaku usaha mandiri.

Berdasarkan Tafsir Al-Misbah (Shihab, 2002), salah satu indikasi kebenaran adalah mengajak manusia untuk merenung, karena dalam penciptaan alam semesta, seperti matahari, bulan, dan berbagai gugusan bintang di langit, serta dalam penataan yang cermat dari sistem langit, serta dalam rotasi bumi dan sumbunya yang menyebabkan pergantian malam dan siang, terdapat bukti-bukti keagungan kekuasaan Allah bagi ulul albab, yaitu mereka yang memiliki akal yang jernih. Allah *Subhanahu Wa Ta'ala* menyatakan dalam Al-Qur'an surat Ali 'Imran/3:190.

إِنَّ فِي خَلْقِ السَّمَوَاتِ وَالْأَرْضِ وَاخْتِلَافِ اللَّيْلِ وَالنَّهَارِ لَآيَاتٍ لِأُولَى الْأَلْبَابِ { ١٩٠ }

“*Sesungguhnya dalam penciptaan langit dan bumi serta pergantian malam dan siang terdapat tanda-tanda (kebesaran Allah) bagi orang yang berakal,*” (Qs. Ali Imron/3:190)

Tafsir Al-Misbah dalam surat Ali Imran ayat 190 menegaskan bahwa sebagian ciptaan Allah seperti langit, bumi, serta siklus pergantian malam dan siang adalah bukti kemahabesan Allah. Penafsiran ini mengaitkan bahwa Allah telah mengklasifikasikan segala ciptaan-Nya, termasuk bumi, langit, malam, dan siang.

Setelahnya, penegasan diperkuat dalam ayat berikutnya, yakni pada ayat 191, dengan penafsiran dari M. Quraish Shihab yang menyatakan bahwa Allah tidak menciptakan alam semesta beserta isinya secara sia-sia atau tanpa tujuan yang pasti. Segala keburukan atau kekurangan yang kita alami, lihat, atau dengar merupakan akibat dari perilaku, dosa, dan kekurangan kita sendiri yang bisa mengakibatkan kita tersesat ke dalam siksa neraka. Oleh karena itu, kami bermohon agar dilindungi dari siksa neraka. Allah *Subhanahu Wa Ta'ala* berfirman dalam Al-Qur'an surat Ali 'Imran/3:191.

الَّذِينَ يَذْكُرُونَ اللَّهَ قِيَامًا وَقُعُودًا وَعَلَىٰ جُنُوبِهِمْ وَيَتَفَكَّرُونَ فِي خَلْقِ السَّمَوَاتِ وَالْأَرْضِ رَبَّنَا

مَا خَلَقْتَ هَذَا بَاطِلًا سُبْحَانَكَ قِنَا عَذَابَ النَّارِ ﴿١٩١﴾

“(yaitu) orang-orang yang mengingat Allah sambil berdiri, duduk, atau dalam keadaan berbaring, dan memikirkan tentang penciptaan langit dan bumi (seraya berkata), “Ya Tuhan kami, tidaklah Engkau menciptakan semua ini sia-sia. Mahasuci Engkau. Lindungilah kami dari azab neraka.” (Qs. Ali Imron/3:191)

Penafsiran ayat tersebut, yang diperjelas oleh M. Quraish Shihab, menyiratkan pentingnya mengingat Allah dalam segala posisi, baik berdiri, duduk,

maupun berbaring, sambil merenungi keagungan penciptaan langit dan bumi. Pesan yang disampaikan adalah bahwa setiap ciptaan Allah memiliki tujuan yang mulia dan tak ada yang sia-sia. Dari sinilah dapat disimpulkan bahwa Allah menciptakan langit, bumi, dan segala isinya dengan penuh hikmah, serta telah memberikan contoh klasifikasi yang sempurna terhadap seluruh ciptaan-Nya.

Penelitian yang dilakukan oleh Amalia (2020) memaparkan bahwa pemanfaatan data kelulusan siswa SMK sebagai sumber informasi diharapkan dapat memberikan wawasan mengenai pola kelulusan siswa SMK melalui teknik analisis data. Salah satu pendekatan yang digunakan adalah mengklasifikasikan data lulusan menggunakan metode *Decision Tree* dan *Naive Bayes*. *Decision Tree* adalah algoritma klasifikasi yang digunakan untuk meramalkan respons dari data yang tersedia, sementara *Naive Bayes* merupakan metode klasifikasi yang berbasis pada probabilitas dan statistik untuk meramalkan kemungkinan di masa depan berdasarkan pengalaman masa lalu, yang dikenal sebagai Teorema Bayes (Ginatra *et al.*, 2021).

## **1.2. Pernyataan Masalah**

Seberapa tepat klasifikasi bidang kerja lulusan SMK jika ditinjau dari faktor linearitas?

## **1.3. Tujuan Penelitian**

Mengukur kinerja klasifikasi bidang kerja lulusan SMK ditinjau dari faktor linearitas yang tepat.

### 3.1 Manfaat Penelitian

Adapun manfaat dari penelitian ini adalah:

1. Bagi unit BKK sekolah dapat menjadi referensi untuk tim BKK dalam proses mencarikan industri pasangan
2. Bagi unit Kurikulum sekolah dapat memberikan informasi terhadap kurikulum untuk sistem pembelajaran di sekolah supaya sesuai dengan kebutuhan industri

### 3.2 Ruang Lingkup Penelitian

Dengan mempertimbangkan paparan sebelumnya dan menetapkan batasan yang tegas dalam penelitian ini, lingkup penelitian ini terbatas pada:

1. Nilai Uji Kompetensi Keahlian (UKK)
2. Nilai Rapot Mata Pelajaran Komputer Jaringan Dasar (KOMJARDAS)
3. Nilai Rapot Mata Pelajaran Pemrograman Dasar (PROGDAS)
4. Nilai Rapot Mata Pelajaran Simulasi Digital (SIMDIG)
5. Nilai Rapot Mata Pelajaran Sistem Komputer (SISKOM)
6. Data siswa dan data alumni yang digunakan hanya untuk bidang *Information Technology (IT)*.
7. Data alumni yang digunakan hanya 3 tahun lulusan terakhir yaitu lulusan tahun 2021, lulusan tahun 2022 dan lulusan tahun 2023.

## BAB II KAJIAN PUSTAKA

### 2.1 Sistem Klasifikasi Serapan Lulusan

Penelitian yang dilakukan oleh Kusuma (2019) menyimpulkan bahwa kemungkinan seorang siswa yang lulus untuk mendapatkan pekerjaan dipengaruhi oleh berbagai faktor, termasuk nilai rata-rata rapor, nilai rata-rata Ujian Nasional, dan nilai Ujian Kompetensi Kejuruan (yang mencakup prakerin, disiplin, tanggung jawab, keterampilan, sikap, dan komunikasi) yang diperlukan dalam proses rekrutmen perusahaan. Dalam penelitiannya, algoritma *Naive Bayes* digunakan untuk menganalisis pola dari variabel-variabel tersebut. Pengujian dilakukan menggunakan *Confusion Matrix* dan kurva ROC. Hasil akurasi penelitian menunjukkan bahwa *Naive Bayes* memberikan tingkat akurasi sebesar 98% dengan nilai AUC sebesar 0.980.

Penelitian yang dilakukan oleh Rifai *et al.* (2019) memproyeksikan tingkat kelulusan peserta uji sertifikasi *Microsoft Office Specialist versi 2013 (Word dan Excel)*. Data yang digunakan mencakup nama peserta dan lima kriteria, yaitu: nama, kriteria 1, kriteria 2, kriteria 3, kriteria 4, kriteria 5 beserta kategorinya.

**Tabel 2.1 Kriteria**

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

Metode *Naive Bayes* digunakan untuk klasifikasi status kelulusan sertifikasi, memungkinkan peserta untuk mengetahui informasi mengenai status lulus dan tidak lulus. Pengembangan sistem ini menerapkan metode *CRISP-DM* guna memastikan proses pengembangan berjalan terstruktur, sementara pengujian dilakukan dengan metode *blackbox* untuk menguji setiap fungsi yang terdapat dalam aplikasi yang sedang dikembangkan. Hasil penelitian ini menunjukkan probabilitas sebesar 0.001042 dengan tingkat akurasi mencapai 99%.

Penelitian yang dilakukan oleh Ula *et al.* (2022) bertujuan untuk menganalisis kesesuaian bidang pekerjaan alumni dengan kompetensi lulusan, lama waktu yang dibutuhkan untuk mendapatkan pekerjaan pertama, serta rata-rata penghasilan yang diperoleh alumni pada pekerjaan awal dan pekerjaan saat ini. Data dikumpulkan melalui penyebaran kuesioner, wawancara telepon, dan aplikasi *WhatsApp* kepada alumni lulusan periode 2012 hingga 2016 dengan total responden sebanyak 117 orang. Pekerjaan diklasifikasikan menjadi dua kategori: Pendidikan (termasuk Guru SMK, Guru SD, dan Guru Bimbel) dan Non-Pendidikan (meliputi *Engineer, Estimator, Quantity Surveyor, Quality Control, Design Engineering, Teller Customer*, dan *Wirausaha*). Metode penelitian yang digunakan adalah kualitatif. Hasil penelitian menunjukkan bahwa rata-rata waktu tunggu alumni sebelum mendapatkan pekerjaan pertama kurang dari 3 bulan dengan persentase 74,4%. Selain itu, persentase alumni yang bekerja sesuai dengan kompetensi lulusan dalam bidang pendidikan adalah 7,70%, sementara dalam bidang non-pendidikan mencapai 92,30%.

Penelitian yang dilakukan oleh Sinaga *et al.* (2021) bertujuan untuk menganalisis tingkat kelulusan siswa dengan menerapkan teknik *data mining* berdasarkan nilai Ujian Akhir Sekolah (UAS), Ujian Nasional (UN), dan nilai akhir (NA) selama tiga tahun terakhir, yaitu dari tahun 2017 hingga 2019. Penelitian ini menyoroti efektivitas metode *Naive Bayes* dalam mengatasi permasalahan informasi yang terintegrasi. Metode penelitian ini melibatkan observasi dan wawancara dengan narasumber dari Sekolah SMK Putra Jaya Batam. Dalam penelitian ini, data dari 70 siswa digunakan sebagai sampel, menghasilkan total 210 data, dengan 167 siswa lulus dan 43 siswa tidak lulus.

Penelitian yang dilakukan oleh Rizmayanti *et al.* (2021) membahas penerapan *data mining* dalam meramalkan kompetensi siswa menggunakan metode *Decision Tree*. Penelitian ini bertujuan untuk meramalkan prestasi belajar siswa berdasarkan data nilai rapor dari empat semester. Data yang digunakan berasal dari siswa tingkat X hingga XI dari berbagai jurusan. Pendekatan ini membantu dalam menilai kemampuan siswa secara akurat dan memberikan wawasan yang sesuai dengan kebutuhan pendidikan di Indonesia. Dengan pola dan strategi pembelajaran yang berfokus pada kemampuan berpikir logis siswa, diharapkan siswa mampu menganalisis permasalahan dengan cermat dan menemukan solusi yang tepat. Namun, tantangan terletak pada ketidaksesuaian antara pendekatan kemampuan berpikir logis siswa dengan kurikulum pendidikan yang tersedia. Selain itu, ketidakmerataan kompetensi guru di berbagai sekolah serta kurangnya dukungan pemerintah turut menjadi hambatan dalam mewujudkan sistem pendidikan yang berbasis pada kemampuan berpikir logis siswa.

Penelitian yang dilakukan oleh Fakhrun Shiddieq *et al.* (2019) mengulas klasifikasi bidang pekerjaan alumni dengan tujuan mengevaluasi tingkat kesesuaian informasi pada setiap kompetensi keahlian dengan jabatan yang diemban oleh para lulusan. Pertanyaan pokok adalah sejauh mana posisi yang diduduki oleh lulusan sesuai dengan keahlian yang dimiliki. Metode Algoritma *K-Means* digunakan untuk melakukan klasifikasi bidang alumni, membedakan antara lulusan yang bekerja sangat sesuai, cukup sesuai, atau tidak sesuai dengan kompetensi keahlian yang dimiliki.

Penelitian yang dilakukan oleh Prasojo *et al.* (2021) membahas prediksi kinerja siswa dalam memasuki pasar kerja berdasarkan nilai mata pelajaran umum, nilai kejuruan, dan nilai sikap, dengan menerapkan metode *data mining Linear Regression, Neural Network, dan Support Vector Machine (SVM)*. Subjek penelitian terdiri dari 101 siswa kelas XII SMK Negeri 1 Kotaagung Timur Tanggamus Lampung. Hasil uji menunjukkan tingkat akurasi dari ketiga metode sebagai berikut: *Neural Network* memiliki tingkat prediksi siswa layak kerja sebesar 77,4% dan siswa tidak layak kerja sebesar 22,6%. *Linear Regression* menunjukkan nilai prediksi siswa layak kerja sebesar 56,9% dan siswa tidak layak kerja sebesar 43,1%. Sedangkan *Support Vector Machine (SVM)* menunjukkan prediksi siswa layak kerja sebesar 72,8% dan siswa tidak layak kerja sebesar 27,2%, dengan performa akurasi sebesar 83,3%. Dalam pengolahan data menggunakan algoritma *Linear Regression*, rumus *Linear Regression* dapat dirumuskan sebagai berikut:  $Y = -0,053 \times \text{Nilai Sikap} + 0,003 \times \text{Nilai Keterampilan} - 0,007 \times \text{Nilai US} - 0,011 \times \text{Nilai Akhir} + 6,860$ .

Penelitian yang dilakukan oleh Kartini *et al.* (2016) mengkaji klasifikasi kelulusan siswa Sekolah Kepolisian Negara Regional Kalimantan Selatan dengan memanfaatkan data master siswa dan rekapitulasi data nilai tahun 2014. Algoritma yang diterapkan dalam proses kelulusan adalah algoritma *K-Mean*. Analisis tersebut mengelompokkan lulusan ke dalam kategori berbeda. Kelompok 1 memiliki rekapitulasi nilai lulusan sedang, kelompok 2 memiliki rekapitulasi nilai lulusan tinggi, dan kelompok 3 memiliki rekapitulasi nilai lulusan rendah. Kelompok 1 menunjukkan nilai cluster pada pusat data intelek sebesar 72,93, nilai kepribadian 71,52, dan nilai kesehatan jasmani 62,63 dengan proporsi 45%. Kelompok 2 menunjukkan nilai cluster pada pusat data intelek sebesar 73,33, nilai kepribadian 73,79, dan nilai kesehatan jasmani 70,80 dengan proporsi 35%. Sementara Kelompok 3 menunjukkan nilai cluster pada pusat data intelek sebesar 72,92, nilai kepribadian 69,95, dan nilai kesehatan jasmani 53,64 dengan proporsi 20%.

Penelitian yang dilakukan oleh Purnama *et al.* (2021) menguji analisis klasifikasi waktu tunggu kerja untuk menilai kemudahan alumni dalam memperoleh pekerjaan dengan menggunakan metode klasifikasi *Support Vector Machines (SVM)* dan *Backpropagation Neural Network (BPNN)*. Kedua metode klasifikasi, baik *BPNN* maupun *SVM* dengan fungsi *Kernel Anova*, mampu menggambarkan klasifikasi data *tracer study* berdasarkan tingkat kemudahan alumni dalam mendapatkan pekerjaan (lancar dan tidak lancar) dengan tingkat akurasi yang hampir sama, yakni sebesar 83,33% untuk *BPNN* dan 83,00% untuk *SVM*. Dengan pemahaman terhadap faktor yang mempengaruhi tingkat kemudahan

dalam memperoleh pekerjaan, diharapkan universitas dapat merumuskan kebijakan yang tepat guna meningkatkan kualitas lulusan.

Penelitian yang dilakukan oleh Janu dkk. (2021) menyoroti pentingnya tingkat akurasi data dalam kehidupan sehari-hari, mengingat kemajuan teknologi informasi yang terus berkembang. Transformasi data menjadi informasi yang bermakna dapat dicapai melalui pemanfaatan sistem *data mining*. Dalam konteks prediksi kelulusan, algoritma yang umum digunakan adalah *Naive Bayes* dan C.45. Penelitian ini bertujuan untuk membandingkan kedua algoritma tersebut dalam hal akurasi prediksi kelulusan. Metode penelitian melibatkan analisis literatur dari berbagai sumber yang relevan serta pemahaman mendalam terhadap data yang berkaitan dengan klasifikasi menggunakan algoritma *Naive Bayes* dan C.45 dalam konteks sistem *data mining*. Hasil penelitian menunjukkan bahwa penggunaan algoritma *Naive Bayes* memberikan tingkat akurasi yang lebih tinggi dibandingkan dengan algoritma C.45.

Ditinjau dari perspektif Islam, terdapat klasifikasi ayat berdasarkan muhkam dan mutasyabih. Allah *Subhanahu Wa Ta'ala* menyatakan dalam Al-Qur'an, Surat Ali Imran/3:7:

هُوَ الَّذِي أَنْزَلَ عَلَيْكَ الْكِتَابَ مِنْهُ آيَاتٌ مُحْكَمَاتٌ هُنَّ أُمُّ الْكِتَابِ وَأُخْرُ مُتَشَابِهَاتٌ فَأَمَّا الَّذِينَ فِي قُلُوبِهِمْ زَيْغٌ فَيَتَّبِعُونَ مَا تَشَابَهَ مِنْهُ ابْتِغَاءَ الْفِتْنَةِ وَابْتِغَاءَ تَأْوِيلِهِ وَمَا يَعْلَمُ تَأْوِيلَهُ إِلَّا اللَّهُ وَالرَّاسِخُونَ فِي الْعِلْمِ يَقُولُونَ آمَنَّا بِهِ كُلٌّ مِّنْ عِنْدِ رَبِّنَا وَمَا يَذَّكَّرُ إِلَّا أُولُو الْأَلْبَابِ

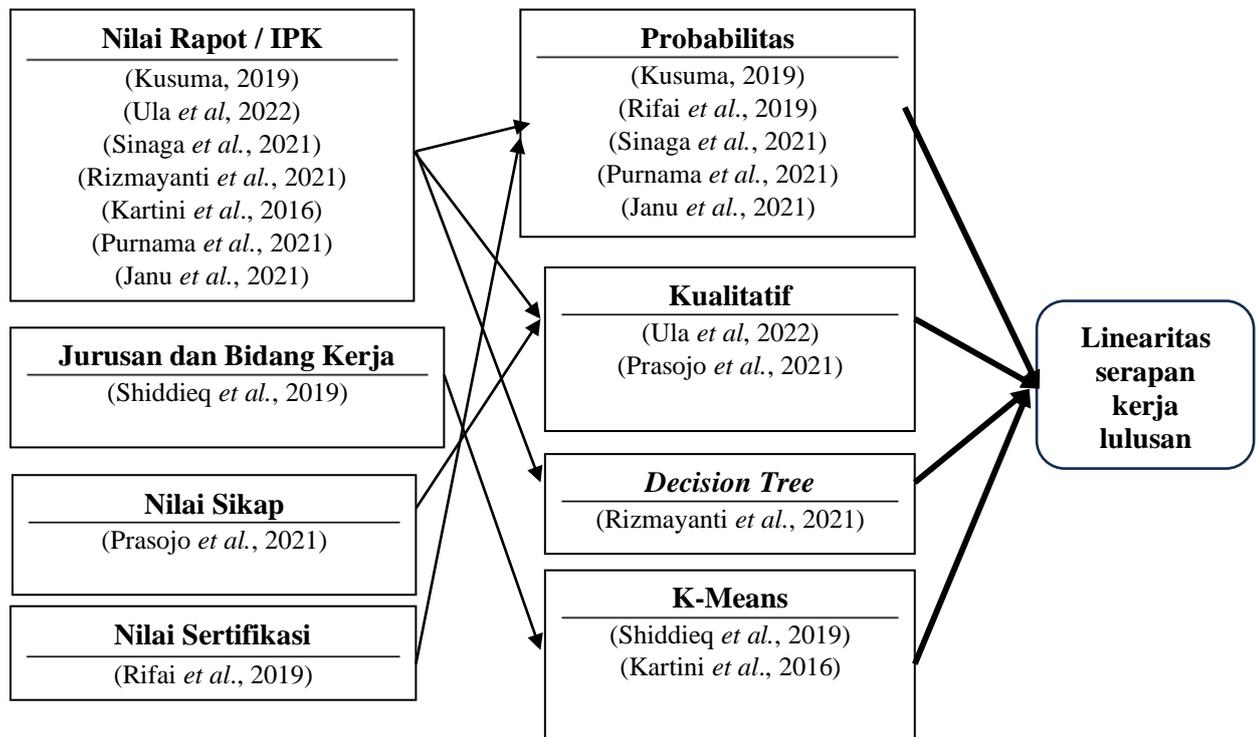
“Dialah yang menurunkan Al Kitab (Al Quran) kepada kamu. Di antara (isi)nya ada ayat-ayat yang muhkamaat, itulah pokok-pokok isi Al qur'an dan yang lain (ayat-ayat) mutasyabihaat. Adapun orang-orang yang dalam hatinya condong kepada kesesatan, maka mereka mengikuti sebahagian ayat-ayat yang

*mutasyabihaat daripadanya untuk menimbulkan fitnah untuk mencari-cari ta'wilnya, padahal tidak ada yang mengetahui ta'wilnya melainkan Allah. Dan orang-orang yang mendalam ilmunya berkata: "Kami beriman kepada ayat-ayat yang mutasyabihaat, semuanya itu dari sisi Tuhan kami". Dan tidak dapat mengambil pelajaran (daripadanya) melainkan orang-orang yang berakal." (QS. Ali-Imran [3:7])*

Menurut Syaiful Arief dalam bukunya "Ulumul Qur'an Untuk Pemula", konsep muhkam dalam terminologi berarti suatu lafadz yang memiliki arti yang jelas dan mampu berdiri sendiri tanpa perlu ditafsirkan karena penyusunannya yang teratur dan tepat, serta maknanya mudah dipahami dan masuk akal. Sementara itu, mutasyabih berasal dari kata "*tasyabaha*" yang berarti serupa, mirip, atau menyerupai yang lainnya (Arief, 2022). Melalui ayat tersebut bisa dianalogikan bahwa dalam pandangan Islam, Allah *Subhanahu Wa Ta'ala* sudah menggunakan klasifikasi untuk membedakan ayat suci Al Qur'an. Seperti yang tertulis pada Al Qur'an pada surat Ali-Imran ayat 7, yang menyampaikan bahwa ayat Al Qur'an telah dibagi menjadi 2 bagian yaitu muhkam dan mutasyabih. Sesuai dengan penelitian ini, akan mengklasifikasikan serapan lulusan SMK berdasarkan linearitas.

## **2.2 Kerangka Teori**

Penelitian ini didukung oleh kerangka teori yang merujuk pada literatur ilmiah dan dijelaskan secara terperinci dalam Gambar 2.1 yang membahas kerangka teori. Sebanyak 10 penelitian sebelumnya telah dilakukan yang memiliki kesamaan. Dalam penelitian sebelumnya, terdapat empat metode yang digunakan, yaitu probabilitas, kualitatif, Pohon Keputusan, dan *K-Means* untuk melakukan klasifikasi.



**Gambar 2.1 Kerangka Teori**

Dalam Gambar 2.1, diuraikan langkah-langkah yang terlibat dalam menyusun studi pustaka sebagai berikut:

Dalam penelitian yang dilakukan oleh Prasojo *et al.* (2021), disampaikan bahwa beberapa metode dan teknik yang digunakan dalam klasifikasi dan estimasi juga dapat diterapkan untuk keperluan prediksi yang sesuai. Mereka menggunakan pendekatan ini untuk meramalkan kinerja siswa SMK setelah memasuki dunia kerja.

Klasifikasi adalah proses pengelompokan data berdasarkan kategori yang relevan guna mempermudah pencarian, penyimpanan, dan analisis data (Fortino, 2023). Klasifikasi data secara otomatis menggunakan algoritma *machine learning* untuk mengelompokkan data yang tidak terstruktur menggunakan *tag* yang telah

ditentukan sebelumnya. Setelah skema kategorisasi ditemukan, logika untuk menentukan kategori tertentu menjadi model untuk mengelompokkan data masa depan yang serupa yang dimasukkan ke dalam *dataset*.

Metode *Naive Bayes* adalah suatu pendekatan yang prinsipil untuk mengintegrasikan informasi eksternal ke dalam proses analisis data (Kantardzic, 2020). Proses ini dimulai dengan distribusi probabilitas yang telah ditentukan untuk kumpulan data yang sedang dianalisis. Karena distribusi ini telah ditetapkan sebelum data yang sebenarnya dipertimbangkan, disebut sebagai distribusi sebelumnya. Data baru akan memperbarui distribusi sebelumnya menjadi distribusi posterior. *Teorema Bayes* menjadi dasar dalam proses pembaruan ini. *Teorema Bayes* memberikan landasan teoritis bagi pendekatan statistik terhadap masalah inferensi klasifikasi.

Algoritma *Naive Bayesian* merupakan metode yang menggunakan pendekatan probabilistik untuk menyelesaikan permasalahan (Kotu dkk., 2015). Pendekatan ini mengestimasi probabilitas untuk setiap nilai variabel kelas terhadap nilai variabel input yang diberikan. Dengan memanfaatkan probabilitas bersyarat, model ini mampu menghitung probabilitas untuk seluruh nilai kelas target pada data yang tidak diketahui, kemudian mengklasifikasikan hasil dengan memilih pemenang berdasarkan probabilitas yang dihasilkan.

Dalam penelitian yang dilakukan oleh Purnama *et al.* (2021), klasifikasi merupakan komponen penting dalam *machine learning* di mana klasifikasi merujuk pada proses pembelajaran fungsi target yang mampu menetapkan setiap atribut ke dalam kelas yang telah ditetapkan sebelumnya. Proses klasifikasi melibatkan

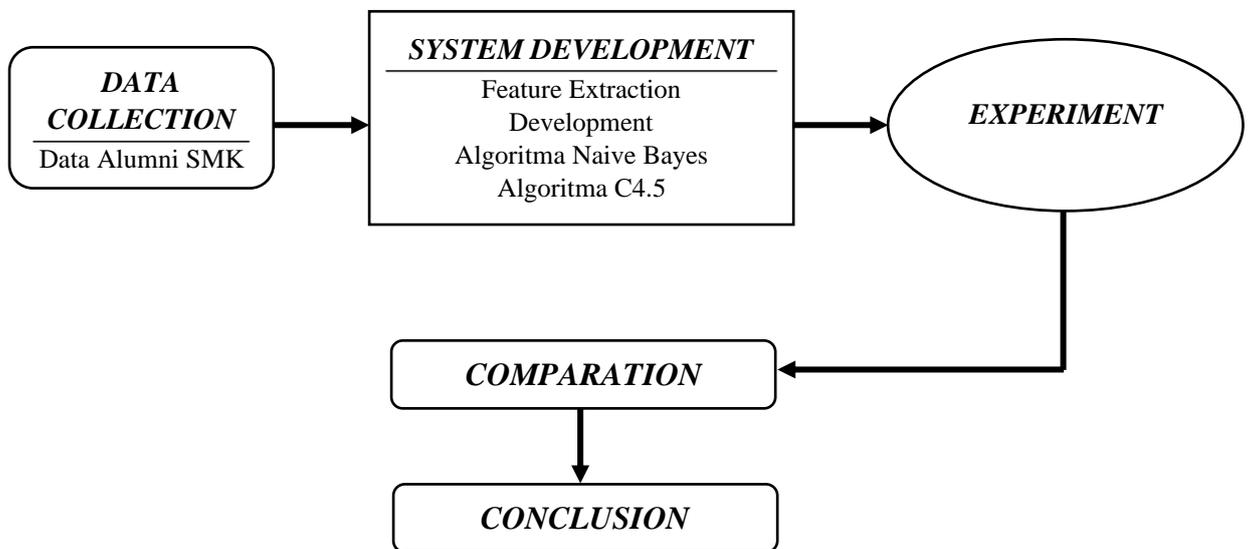
tahapan *training* dan *testing*. *Training* melibatkan penggunaan data *training*, sementara *testing* melibatkan evaluasi model yang dihasilkan dari *training* menggunakan data *testing*.

Berdasarkan hasil kajian pustaka dan kerangka teori, peneliti memilih menggunakan perhitungan dengan probabilitas untuk menyelesaikan masalah linearitas serapan alumni.

## BAB III DESAIN PENELITIAN

### 3.1 Prosedur Penelitian

Penelitian ini memiliki desain yang akan digunakan dan tergambar jelas pada Gambar 3.1. tentang desain penelitian



**Gambar 3.1 Desain Penelitian**

Prosedur penelitian yang ditampilkan pada Gambar 3.1 menjelaskan desain rancangan penelitian yang dimulai dari *data collection*, *feature extraction development*, *system development (Naive Bayes dan C4.5)*, *experiment*, *comparison* algoritma *Naive Bayes* dan *C4.5*, dan *conclusion*.

#### 3.1.1 Data Collection

Pengumpulan data merupakan fase yang krusial dalam suatu penelitian. Ketersediaan data memegang peranan penting dalam proses pengolahan dan analisis selanjutnya. Proses pengumpulan data harus dikerjakan dengan teknik yang

akurat dan menjamin keabsahan data yang diperoleh. Pengumpulan data yang dilakukan secara konseptual ditunjang dengan studi literatur, jurnal penelitian, dan sumber informasi terkait yang mempertajam pemahaman terhadap permasalahan yang sedang diteliti. Penelitian ini menggunakan data primer yang bersumber dari catatan alumni tiga tahun terakhir (2021, 2022, dan 2023) dari SMK Telkom Malang. Proses pengambilan data dilakukan melalui aplikasi alumni Tracert yang merupakan bagian dari sistem informasi sekolah, sementara data nilai diperoleh dari pihak kurikulum sekolah. Sebanyak 1533 data alumni SMK berhasil terkumpul untuk analisis. *Detail* parameter yang digunakan dapat ditemukan pada Tabel 3.1 yang memuat informasi tentang parameter yang relevan.

**Tabel 3.1 Parameter**

<b>Parameter</b>	<b>Kriteria</b>
Jurusan	RPL
	TKJ
Tahun Lulusan	2021
	2022
	2023
Nilai UKK	Nilai > 90
	Nilai = 80 – 89
	Nilai = 70 – 79
	Nilai < 69
Nilai KOMJARDAS	Nilai > 90
	Nilai = 80 – 89

	Nilai = 70 – 79
	Nilai = 60 – 69
	Nilai < 59
Nilai PROGDAS	Nilai > 90
	Nilai = 80 – 89
	Nilai = 70 – 79
	Nilai = 60 – 69
	Nilai < 59
Nilai SIMDIG	Nilai > 90
	Nilai = 80 – 89
	Nilai = 70 – 79
	Nilai = 60 – 69
	Nilai < 59
Nilai SISKOM	Nilai > 90
	Nilai = 80 – 89
	Nilai = 70 – 79
	Nilai = 60 – 69
	Nilai < 59

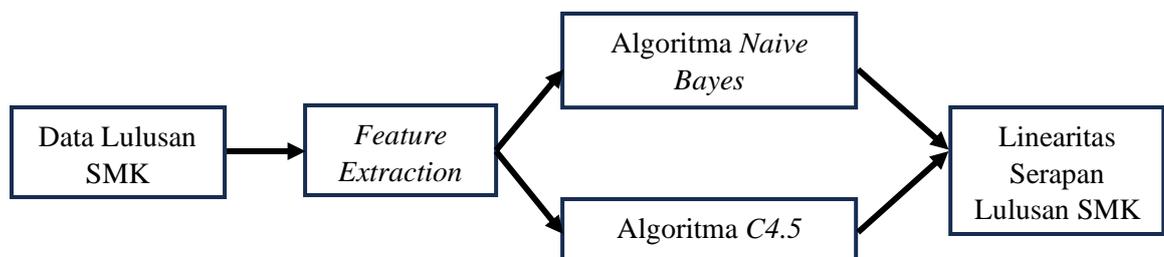
Pemberian parameter pada Tabel 3.1. tentang parameter yang dijadikan variabel untuk mengklasifikasikan liniaritas serapan lulusan SMK. Parameter jurusan diberikan pilihan jawaban yaitu jurusan RPL dengan pemahaman bahwa alumni berada pada kelas RPL sebelum lulus, dan jurusan TKJ dengan pemahaman

bahwa alumni berada pada kelas TKJ sebelum lulus. Parameter tahun lulus diberikan pilihan jawaban yaitu 2021 dengan pemahaman alumni yang lulus pada tahun 2021, untuk parameter 2022 dengan pemahaman alumni yang lulus pada tahun 2022, dan untuk parameter 2023 dengan pemahaman alumni yang lulus pada tahun 2023. Parameter UKK diberikan pilihan jawaban yaitu nilai di atas 90 dengan pemahaman alumni tersebut mendapatkan nilai UKK diatas 90, untuk nilai: 80-89 dengan pemahaman alumni tersebut mendapatkan nilai UKK antara 80-89, untuk nilai: 70-79 dengan pemahaman alumni tersebut mendapatkan nilai UKK antara 70-79, untuk nilai di atas 69 dengan pemahaman alumni tersebut mendapatkan nilai UKK dibawah 69. Parameter KOMJARDAS diberikan pilihan jawaban yaitu nilai di atas 90 dengan pemahaman alumni tersebut mendapatkan nilai KOMJARDAS diatas 90, untuk nilai: 80-89 dengan pemahaman alumni tersebut mendapatkan nilai KOMJARDAS antara 80-89, untuk nilai: 70-79 dengan pemahaman alumni tersebut mendapatkan nilai KOMJARDAS antara 70-79, untuk nilai: 60-69 dengan pemahaman alumni tersebut mendapatkan nilai KOMJARDAS dibawah antara 60-69, untuk nilai di bawah 59 dengan pemahaman alumni tersebut mendapatkan nilai KOMJARDAS dibawah 59. Parameter PROGDAS diberikan pilihan jawaban yaitu nilai di atas 90 dengan pemahaman alumni tersebut mendapatkan nilai PROGDAS diatas 90, untuk nilai: 80-89 dengan pemahaman alumni tersebut mendapatkan nilai PROGDAS antara 80-89, untuk nilai: 70-79 dengan pemahaman alumni tersebut mendapatkan nilai PROGDAS antara 70-79, nilai: 60-69 dengan pemahaman alumni tersebut mendapatkan nilai PROGDAS dibawah antara 60-69, untuk nilai di bawah 59 dengan pemahaman alumni tersebut mendapatkan nilai PROGDAS

dibawah 59. Parameter SIMDIG diberikan pilihan jawaban yaitu nilai di atas 90 dengan pemahaman alumni tersebut mendapatkan nilai SIMDIG diatas 90, untuk nilai: 80-89 dengan pemahaman alumni tersebut mendapatkan nilai SIMDIG antara 80-89, untuk nilai: 70-79 dengan pemahaman alumni tersebut mendapatkan nilai SIMDIG antara 70-79, untuk nilai: 60-69 dengan pemahaman alumni tersebut mendapatkan nilai SIMDIG dibawah antara 60-69, untuk nilai di bawah 59 dengan pemahaman alumni tersebut mendapatkan nilai SIMDIG dibawah 59. Parameter SISKOM diberikan pilihan jawaban yaitu nilai di atas 90 dengan pemahaman alumni tersebut mendapatkan nilai SISKOM diatas 90, untuk nilai: 80-89 dengan pemahaman alumni tersebut mendapatkan nilai SISKOM antara 80-89, untuk nilai: 70-79 dengan pemahaman alumni tersebut mendapatkan nilai SISKOM antara 70-79, untuk nilai: 60-69 dengan pemahaman alumni tersebut mendapatkan nilai SISKOM dibawah antara 60-69, untuk nilai di bawah 59 dengan pemahaman alumni tersebut mendapatkan nilai SISKOM dibawah 59. Penentuan kriteria nilai didasarkan pada kriteria yang sudah dibuat dari pihak sekolah.

### 3.1.2 *System Development*

Penelitian ini memiliki *system development* yang akan digunakan dan tergambar jelas pada Gambar 3.2. tentang desain *system*



**Gambar 3.2 Desain System**

Desain system development yang ditampilkan pada Gambar 3.2 menjelaskan bahwa akan dilakukan proses pengumpulan data lulusan SMK. Selanjutnya akan ditambahkan dengan feature extraction yang berisi Kelas, Tahun Lulus, Bidang, Nilai Uji Kompetensi Kejuruan (UKK), Nilai Komputer Jaringan Dasar (KOMJARDAS), Nilai Simulasi Digital (SIMDIG), Nilai Pemrograman Dasar (PROGDAS), Nilai Sistem Komputer (SISKOM). Kemudian akan diproses dengan algoritma *Naive Bayes* dan algoritma C4.5 sebagai pembanding yang digunakan untuk mendapatkan hasil yaitu linearitas serapan lulusan SMK.

### 3.1.3 Feature Extraction

Tahap feature extraction menentukan *class* klasifikasi diantaranya *class* linier dan *class* tidak linier. Setelah menentukan *class* klasifikasi tahapan berikutnya adalah menentukan tingkatan jawaban dari kriteria pada masing-masing parameter terlihat pada Tabel 3.3. tentang tingkatan kriteria parameter.

**Tabel 3.3 Tingkatan Kriteria Parameter**

Parameter	Kriteria	Tingkatan
Jurusan	RPL	RPL 1
		RPL 2
		RPL 3
		RPL 4
		RPL 5
		RPL 6
		RPL 7
		RPL 8

		TKJ 1
		TKJ 2
		TKJ 3
		TKJ 4
		TKJ 5
		TKJ 6
Tahun Lulus	2021	Lulusan 2021
	2022	Lulusan 2022
	2023	Lulusan 2023
Nilai UKK	Nilai = 90 – 100	Sangat Kompeten
	Nilai = 80 – 89	Kompeten
	Nilai = 70 – 79	Cukup Kompeten
	Nilai < 69	Tidak Kompeten
Nilai KOMJARDAS	Nilai = 90 – 100	Predikat A
	Nilai = 80 – 89	Predikat B
	Nilai = 70 – 79	Predikat C
	Nilai = 60 – 69	Predikat D
	Nilai < 59	Predikat E
Nilai PROGDAS	Nilai = 90 – 100	Predikat A
	Nilai = 80 – 89	Predikat B
	Nilai = 70 – 79	Predikat C
	Nilai = 60 – 69	Predikat D

	Nilai < 59	Predikat E
Nilai SIMDIG	Nilai = 90 – 100	Predikat A
	Nilai = 80 – 89	Predikat B
	Nilai = 70 – 79	Predikat C
	Nilai = 60 – 69	Predikat D
	Nilai < 59	Predikat E
Nilai SISKOM	Nilai = 90 – 100	Predikat A
	Nilai = 80 – 89	Predikat B
	Nilai = 70 – 79	Predikat C
	Nilai = 60 – 69	Predikat D
	Nilai < 59	Predikat E

Pada Tabel 3.3 tentang tingkatan kriteria parameter dijelaskan bahwa ada tingkatan dalam setiap kriteria pada parameter yang sudah dibuat. Untuk parameter jurusan dengan kriteria RPL dan TKJ mempunyai beberapa tingkatan. Untuk kriteria RPL mempunyai 8 tingkatan yaitu tingkatan kriteria RPL 1 dengan pemahaman bahwa alumni berada pada jurusan RPL pada urutan kelas RPL 1, untuk tingkatan kriteria RPL 2 dengan pemahaman bahwa alumni berada pada jurusan RPL pada urutan kelas RPL 2, untuk tingkatan kriteria RPL 3 dengan pemahaman bahwa alumni berada pada jurusan RPL pada urutan kelas RPL 3, untuk tingkatan kriteria RPL 4 dengan pemahaman bahwa alumni berada pada jurusan RPL pada urutan kelas RPL 4, untuk tingkatan kriteria RPL 5 dengan pemahaman bahwa alumni berada pada jurusan RPL pada urutan kelas RPL 5,

untuk tingkatan kriteria RPL 6 dengan pemahaman bahwa alumni berada pada jurusan RPL pada urutan kelas RPL 6, untuk tingkatan kriteria RPL 7 dengan pemahaman bahwa alumni berada pada jurusan RPL pada urutan kelas RPL 7, dan untuk tingkatan kriteria RPL 8 dengan pemahaman bahwa alumni berada pada jurusan RPL pada urutan kelas RPL 8. Pada kriteria TKJ juga memiliki tingkatan kriteria yaitu tingkatan kriteria TKJ 1 dengan pemahaman bahwa alumni berada pada jurusan TKJ pada urutan kelas TKJ 1, untuk tingkatan kriteria TKJ 2 dengan pemahaman bahwa alumni berada pada jurusan TKJ pada urutan kelas TKJ 2, untuk tingkatan kriteria TKJ 3 dengan pemahaman bahwa alumni berada pada jurusan TKJ pada urutan kelas TKJ 3, untuk tingkatan kriteria TKJ 4 dengan pemahaman bahwa alumni berada pada jurusan TKJ pada urutan kelas TKJ 4, untuk tingkatan kriteria TKJ 5 dengan pemahaman bahwa alumni berada pada jurusan TKJ pada urutan kelas TKJ 5, dan untuk tingkatan kriteria TKJ 6 dengan pemahaman bahwa alumni berada pada jurusan TKJ pada urutan kelas TKJ 6. Parameter UKK diberikan pilihan jawaban yaitu kriteria nilai di atas 90 dengan pemahaman alumni tersebut mendapatkan nilai UKK diatas 90 dengan predikat sangat kompeten, untuk kriteria nilai: 80-89 dengan pemahaman alumni tersebut mendapatkan nilai UKK antara 80-89 dengan predikat kompeten, untuk kriteria nilai: 70-79 dengan pemahaman alumni tersebut mendapatkan nilai UKK antara 70-79 dengan predikat cukup kompeten, untuk kriteria nilai di atas 69 dengan pemahaman alumni tersebut mendapatkan nilai UKK dibawah 69 dengan predikat tidak kompeten. Parameter KOMJARDAS diberikan pilihan jawaban yaitu kriteria nilai di atas 90 dengan pemahaman alumni tersebut mendapatkan nilai KOMJARDAS diatas 90 dengan

predikat A, untuk kriteria nilai: 80-89 dengan pemahaman alumni tersebut mendapatkan nilai KOMJARDAS antara 80-89 dengan predikat B, untuk kriteria nilai: 70-79 dengan pemahaman alumni tersebut mendapatkan nilai KOMJARDAS antara 70 -79 dengan predikat C, untuk kriteria nilai: 60-69 dengan pemahaman alumni tersebut mendapatkan nilai KOMJARDAS dibawah antara 60-69 dengan predikat D, untuk kriteria nilai di bawah 59 dengan pemahaman alumni tersebut mendapatkan nilai KOMJARDAS dibawah 59 dengan predikat E. Parameter PROGDAS diberikan pilihan jawaban yaitu kriteria nilai di atas 90 dengan pemahaman alumni tersebut mendapatkan nilai PROGDAS diatas 90 dengan predikat A, untuk kriteria nilai: 80-89 dengan pemahaman alumni tersebut mendapatkan nilai PROGDAS antara 80-89 dengan predikat B, untuk kriteria nilai: 70-79 dengan pemahaman alumni tersebut mendapatkan nilai PROGDAS antara 70-79 dengan predikat C, untuk kriteria nilai: 60-69 dengan pemahaman alumni tersebut mendapatkan nilai PROGDAS dibawah antara 60-69 dengan predikat D, untuk kriteria nilai di bawah 59 dengan pemahaman alumni tersebut mendapatkan nilai PROGDAS dibawah 59 dengan predikat E. Parameter SIMDIG diberikan pilihan jawaban yaitu kriteria nilai di atas 90 dengan pemahaman alumni tersebut mendapatkan nilai SIMDIG diatas 90 dengan predikat A, untuk kriteria nilai: 80-89 dengan pemahaman alumni tersebut mendapatkan nilai SIMDIG antara 80-89 dengan predikat B, untuk kriteria nilai: 70-79 dengan pemahaman alumni tersebut mendapatkan nilai SIMDIG antara 70-79 dengan predikat C, untuk kriteria nilai: 60-69 dengan pemahaman alumni tersebut mendapatkan nilai SIMDIG dibawah antara 60-69 dengan predikat D, untuk kriteria nilai di bawah 59 dengan

pemahaman alumni tersebut mendapatkan nilai SIMDIG dibawah 59 dengan predikat E. Parameter SISKOM diberikan pilihan jawaban yaitu kriteria nilai di atas 90 dengan pemahaman alumni tersebut mendapatkan nilai SISKOM diatas 90 dengan predikat A, untuk kriteria nilai: 80-89 dengan pemahaman alumni tersebut mendapatkan nilai SISKOM antara 80-89 dengan predikat B, untuk kriteria nilai: 70-79 dengan pemahaman alumni tersebut mendapatkan nilai SISKOM antara 70 - 79 dengan predikat C, untuk kriteria nilai: 60-69 dengan pemahaman alumni tersebut mendapatkan nilai SISKOM dibawah antara 60-69 dengan predikat D, untuk kriteria nilai di bawah 59 dengan pemahaman alumni tersebut mendapatkan nilai SISKOM dibawah 59 dengan predikat E.

Pada tahap berikutnya, langkah pertama adalah membentuk basis kasus yang akan mencakup 1533 kasus. Setelah itu, langkah selanjutnya adalah melakukan pencocokan menggunakan algoritma *Naive Bayes*. *Dataset* Rician yang digunakan untuk proses klasifikasi dapat ditemukan dalam Tabel 3.4 yang menggambarkan data alumni.

**Tabel 3.4 Dataset Alumni**

NO	Nama	Jurusan	Tahun Lulus	U K K	KOM JAR DAS	PROG DAS	SIM DIG	SIS KOM	Industri	Bidang
1	ACHMAD AZHAR ALDIANSYAH	RPL 1	2023	85	77	81	83	83	PT TELKOM AKSES KEDIRI	Teknologi
2	ADINDA KHARISMA PUTRI TANDAWIJAYA	RPL 1	2023	76	87	84	85	84	CV SINARA SUKSES BERSAMA	Teknologi
3	AIZAR RAHIMA SUPRAYITNO	RPL 1	2023	95	91	91	90	89	BINA NUSANTA RA UNIVERSI TY	Sistem Informasi

4	AMELIA INDRIANA	RPL 1	2023	85	86	84	84	82	PT FINNET INDONESIA A	Teknologi
5	ARDHIAN CALWA NUGRAHA	RPL 1	2023	98	86	86	86	86	TELKOM UNIVERSI TY BANDUN G	TEKNIK INFORMA TIKA
6	ARIESTA PURWA NUGRAHA	RPL 1	2023	82	78	79	80	80	Politeknik Elektronika Negeri Surabaya	TEKNIK INFORMA TIKA
7	AULIYA PUTRI	RPL 1	2023	93	87	87	84	82	Politeknik Negeri Malang	Teknik Jaringan Telekomuni kasi Digital
8	AYU AISYAH FATIMATUH SYAHRA	RPL 1	2023	89	85	85	87	86	PT ANGKAS A PURA I JUANDA	Teknologi
9	BASHAR LENTERA JIHAD KHAWARIQ	RPL 1	2023	88	83	87	82	82	SEVEN INC	Teknologi
10	BINTANG ILHAM PRATAMA	RPL 1	2023	95	86	86	82	82	TELKOM UNIVERSI TY BANDUN G	Teknik Komputer
11	BRYAN BECKHAM WIBISONO EFENDI	RPL 1	2023	93	85	77	80	80	PT ARTHA ELEKTRO NIK INDONESIA A OFFICE "EZY PAY"	Teknologi
12	CESSALYA DEA IVANA PUTRI	RPL 1	2023	84	87	86	85	83	Insura	Teknologi
13	CHASIDIYAH CHESARYA AL ASYHARI	RPL 1	2023	77	82	85	87	87	PT ANGKAS A PURA I JUANDA	Teknologi
14	CHEISYA VALDA WIBAWANINGRU M	RPL 1	2023	92	90	88	88	86	TELKOM UNIVERSI TY BANDUN G	Sistem Informasi
15	CSYA ARUMAE	RPL 1	2023	54	84	83	84	82	PT ARTHA ELEKTRO NIK INDONESIA A OFFICE "EZY PAY"	Teknologi
16	DAFA ADHAM JUSTITIA	XII RPL 1	2023	88	79	82	82	82	PDAM DELTA TIRTA KABUPAT EN SIDOARJO	Teknologi

17	DAVINA ARAFA AURELIA	XII RPL 1	2023	82	85	91	86	88	PT THUNDER LABS INDONESIA	Teknologi
18	DEASY NABILA NUR SAFITRI	XII RPL 1	2023	88	86	86	82	82	UNIVERSI TAS BRAWIJA YA (FAK EKONOMI & BISNIS)	Teknologi
19	DEAZARD MUHAMMAD ARRAYYAN	XII RPL 1	2023	93	77	80	82	83	DINAS PERHUBU NGAN KOTA MALANG	Teknologi
20	DEBI ANGGI PERMATA	XII RPL 1	2023	53	86	93	94	94	Universitas Negeri Malang	Teknologi Rekayasa Sistem Elektronika
...	.....	...	...	...	...	...	...	...	.....	.....
1529	SYAHROTUL FITRIYAH	XII TKJ 6	2021	87	93	88	85	88	Institut Teknologi Sepuluh Nopember	Sistem Informasi
1530	TANAYA WIDYA DIANINGATI	XII TKJ 6	2021	74	93	96	85	90	PT JALA DIGITAL INDONESIA	Teknologi
1531	TAWANG NARPENDYASIWI	TKJ 6	2021	60	86	93	86	85	Universitas Terbuka	Manajemen
1532	TRY PRASETYO ADI	TKJ 6	2021	88	82	92	84	82	DINAS KOMUNIK ASI DAN INFORMA TIKA KAB. MALANG	Teknologi
1533	VITORIO RISWANDANA	TKJ 6	2021	80	81	91	85	79	IMPRO DIGITAL (CV RAJAKA MULTIJAS A)	Teknologi

Penentuan *dataset* pada Tabel 3.4 tentang data alumni didapat dari data alumni *tracert* dan digabungkan dengan data nilai dari alumni SMK Telkom Malang. Dengan parameter yang sudah ditentukan sebelumnya.

Setelah mendapat data alumni, akan diproses untuk *data cleaning* yang akan digunakan untuk *data training* dan *data testing* yang ditampilkan pada Tabel 3.5 tentang *dataset* alumni *cleaning*

**Tabel 3.5 Dataset Alumni Cleaning**

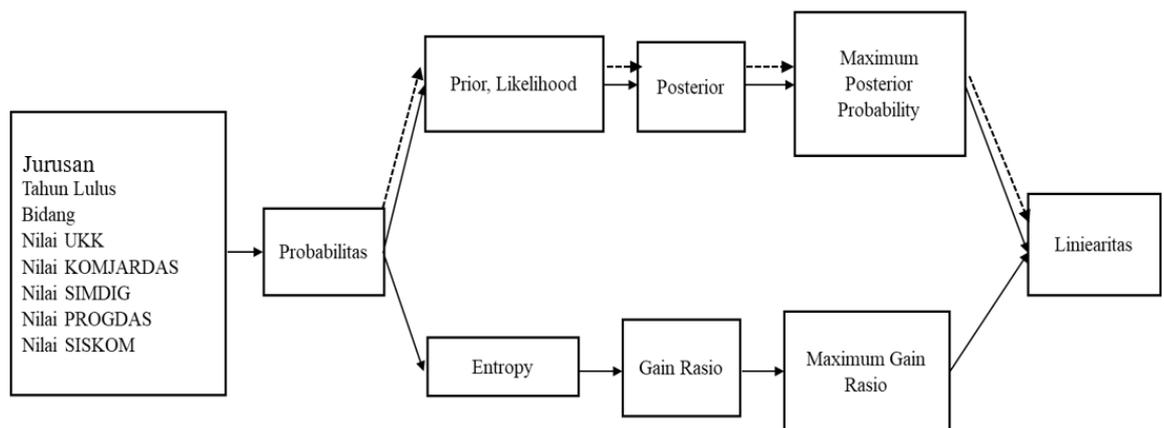
NO	Nama	Jurusan	Tahun Lulus	UKK	KOM JAR DAS	PROG DAS	SIM DIG	SIS KOM	Industri	Bidang
1	ACHMAD AZHAR ALDIANSYAH	RPL 1	2023	K	C	B	B	B	PT TELKOM AKSES KEDIRI	Teknologi
2	ADINDA KHARISMA PUTRI TANDAWIJAYA	RPL 1	2023	CK	B	B	B	B	CV SINARA SUKSES BERSAMA	Teknologi
3	AIZAR RAHIMA SUPRAYITNO	RPL 1	2023	SK	A	A	A	B	BINA NUSANTARA UNIVERSITY	Sistem Informasi
4	AMELIA INDRIANA	RPL 1	2023	K	B	B	B	B	PT FINNET INDONESIA	Teknologi
5	ARDHIAN CALWA NUGRAHA	RPL 1	2023	SK	B	B	B	B	TELKOM UNIVERSITY BANDUNG	TEKNIK INFORMATIKA
6	ARIESTA PURWA NUGRAHA	RPL 1	2023	K	C	C	B	B	Politeknik Elektronika Negeri Surabaya	TEKNIK INFORMATIKA
7	AULIYA PUTRI	RPL 1	2023	SK	B	B	B	B	Politeknik Negeri Malang	Teknik Jaringan Telekomunikasi Digital
8	AYU AISYAH FATIMATUH SYAHRA	RPL 1	2023	K	B	B	B	B	PT ANGKASA PURA IJUANDA	Teknologi
9	BASHAR LENTERA JIHAD KHAWARIQ	RPL 1	2023	K	B	B	B	B	SEVEN INC	Teknologi
10	BINTANG ILHAM PRATAMA	RPL 1	2023	SK	B	B	B	B	TELKOM UNIVERSITY BANDUNG	Teknik Komputer
11	BRYAN BECKHAM WIBISONO EFENDI	RPL 1	2023	SK	B	C	B	B	PT ARTHA ELEKTRONIK INDONESIA OFFICE "EZYPAY"	Teknologi
12	CESSALYA DEA IVANA PUTRI	RPL 1	2023	K	B	B	B	B	Insura	Teknologi

13	CHASIDIYAH CHESARYA AL ASYHARI	RPL 1	2023	CK	B	B	B	B	PT ANGKAS A PURA I JUANDA	Teknologi
14	CHEISYA VALDA WIBAWANINGRUM	RPL 1	2023	SK	A	B	B	B	TELKOM UNIVERSI TY BANDUN G	Sistem Informasi
15	CSYA ARUMAE	RPL 1	2023	TK	B	B	B	B	PT ARTHA ELEKTRO NIK INDONESIA OFFICE "EZYPA"	Teknologi
16	DAFA ADHAM JUSTITIA	XII RPL 1	2023	K	C	B	B	B	PDAM DELTA TIRTA KABUPAT EN SIDOARJO	Teknologi
17	DAVINA ARAFA AURELIA	XII RPL 1	2023	K	B	A	B	B	PT THUNDER LABS INDONESIA	Teknologi
18	DEASY NABILA NUR SAFITRI	XII RPL 1	2023	K	B	B	B	B	UNIVERSI TAS BRAWIJA YA (FAK EKONOMI & BISNIS)	Teknologi
19	DEAZARD MUHAMMAD ARRAYYAN	XII RPL 1	2023	SK	C	B	B	B	DINAS PERHUBU NGAN KOTA MALANG	Teknologi
20	DEBI ANGGI PERMATA	XII RPL 1	2023	TK	B	A	A	A	Universitas Negeri Malang	Teknologi Rekayasa Sistem Elektronika
...	.....	...	...	...	...	...	...	...	.....	.....
1529	SYAHROTUL FITRIYAH	XII TKJ 6	2021	K	A	B	B	B	Institut Teknologi Sepuluh Nopember	Sistem Informasi
1530	TANAYA WIDYA DIANINGATI	XII TKJ 6	2021	CK	A	A	B	A	PT JALA DIGITAL INDONESIA	Teknologi
1531	TAWANG NARPENDYASIWI	TKJ 6	2021	TK	B	A	B	B	Universitas Terbuka	Manajemen
1532	TRY PRASETYO ADI	TKJ 6	2021	K	B	A	B	B	DINAS KOMUNIK ASI DAN INFORMA TIKA	Teknologi

									KAB. MALANG	
1533	VITORIO RISWANDANA	TKJ 6	2021	K	B	A	B	C	IMPRO DIGITAL (CV RAJAKA MULTIJAS A)	Teknologi

#### 4.1 Konseptual Model

Dalam model konseptual dijelaskan secara rinci proses pengolahan *dataset* alumni SMK hingga tercapainya hubungan linear sebagaimana tergambar pada Gambar 3.3 mengenai model konseptual.



**Gambar 3.3 Konseptual Model**

Berdasarkan Gambar 3.3 yang menggambarkan model konseptual, atribut atau parameter yang dipergunakan akan dikenai kriteria untuk memudahkan proses pengumpulan *dataset*.

Berikutnya, dilakukan penghitungan probabilitas untuk setiap atribut atau parameter yang digunakan. Setelah probabilitas dari masing-masing parameter diketahui, langkah selanjutnya adalah menerapkan algoritma *Naive Bayes* untuk menghitung nilai prior, evidence, dan likelihood guna menghasilkan nilai posterior.

Nilai posterior tertinggi akan menjadi prioritas dalam proses berikutnya untuk mencapai tingkat kelulusan siswa SMK yang diharapkan.

Langkah selanjutnya adalah menggunakan algoritma *Decision Tree* (C4.5) untuk mengukur tingkat entropy dari setiap atribut atau parameter. Entropy yang dihasilkan kemudian diolah untuk menghitung nilai gain. Nilai gain ini kemudian dinormalisasi dengan membaginya dengan info gain, sehingga menghasilkan rasio gain. Rasio gain tertinggi akan menjadi root utama dan root-root berikutnya. Selanjutnya, sebuah pohon keputusan dibangun berdasarkan semua root yang terbentuk. Pohon keputusan yang dihasilkan akan berperan sebagai model dalam melakukan klasifikasi linearitas lulusan SMK.

## BAB IV

### METODE NAIVE BAYES

#### 4.1 Desain Metode

Peneliti menerapkan metode *Naive Bayes* sebagaimana terlihat pada Gambar 3.2. Tahapan selanjutnya dalam penelitian ini melibatkan pengujian data menggunakan algoritma *Machine Learning Naive Bayes*, yang berdasarkan konsep probabilitas  $P(c|x)$ , yaitu probabilitas kelas  $c$  jika diketahui suatu variable  $x$ . proses yang dilakukan untuk menentukan nilai suatu kelas  $c \in C$  dari suatu dokumen  $x \in X$  dengan  $C = [c_1, c_2, c_3, \dots c_i]$  dan  $X = [x_1, x_2, x_3, \dots x_j]$ . Sedangkan penentuan kelas dalam klasifikasi dokumen dilakukan dengan cara memilih nilai maksimum dari  $p(c|x)$  berdasarkan distribusi probabilitas.

$$P = \{p(c|x) \mid c \in C \text{ dan } x \in X\} \quad (4.1)$$

Sehingga dokumen  $C$  ke I dapat dipresentasikan sebagai vector dan nilai-nilai fitur yang ada pada dokumen tersebut yaitu  $C = [c_1, c_2, c_3, \dots c_i]$ , nilai-nilai dari elemen setiap vector  $X$  adalah nilai untuk fitur pada himpunan fitur atau variable  $X = [x_1, x_2, x_3, \dots x_j]$  dengan  $x_j$  adalah nilai dari fitur  $x$  ke  $j$  pada dokumen elemen  $X$ . Berdasarkan algoritma *Naive Bayes* adalah persamaan perhitungan nilai dari probabilitas  $P(c|x)$  sebagai berikut persamaannya:

$$Posterior = \frac{Likelihood \cdot Prior}{Evidance} \quad (4.2)$$

$$P(c|x) = \frac{P(x|c) P(c)}{P(x)} \quad (4.3)$$

Keterangan:

$P(c|x)$  = Probabilitas hipotesis berdasarkan kondisi posteriori probabilitas

$P(x|c)$  = Probabilitas berdasarkan kondisi  $c$  (*likelihood*)

$P(c)$  = Probabilitas hipotesis prior probabilitas (*prior*)

$P(x)$  = Probabilitas  $x$  (*evidence*)

#### 4.1.1 Perhitungan Nilai Prior

Perhitungan nilai *prior*  $P(c)$  didapatkan dari rumus berikut :

$$P(c_i) = \frac{\sum(c_1+c_2+\dots+c_i)}{\sum data} \quad (4.4)$$

Keterangan :

$P(c_i)$  = nilai *prior*

$\sum(c_1 + c_2 + \dots c_i)$  = Jumlah dari probabilitas fitur atau parameter linier

$\sum data$  = Jumlah dari semua data

Dengan dasar rumus *prior* tersebut, didapatkan dinilai *prior* linier  $P_{(Linier)}$

dan *prior* tidak linier  $P_{(TidakLinier)}$  sebagai berikut :

$$P_{(Linier)} = \frac{\sum Linier}{\sum data} = \frac{1265}{1533} = 0,83$$

$$P_{(TidakLinier)} = \frac{\sum TidakLinier}{\sum data} = \frac{268}{1533} = 0,17$$

Dari hasil perhitungan, nilai *prior* linier  $P_{(Linier)}$  didapatkan nilai 0,83 dan untuk nilai *prior* tidak linier  $P_{(TidakLinier)}$  adalah 0,17.

#### 4.1.2 Perhitungan Nilai Evidence

Perhitungan nilai *evidence*  $P(x)$  didapatkan dari rumus sebagai berikut :

$$P(x_j) = \frac{\sum(x_1+x_2+\dots+x_j)}{\sum data} \quad (4.5)$$

Keterangan :

$P(x_j)$  = nilai *evidence*

$\sum(x_1 + x_2 + \dots x_j)$  = Jumlah dari probabilitas fitur atau parameter linier

$\sum data$  = Jumlah dari semua data

- Perhitungan dinilai *evidence* pada setiap kriteria pada parameter jurusan sebagai berikut :

$$P_{(Rpl)} = \frac{\sum rpl}{\sum data} = \frac{872}{1533} = 0,57$$

$$P_{(Tkj)} = \frac{\sum tkj}{\sum data} = \frac{661}{1533} = 0,43$$

Dari hasil perhitungan, didapatkan nilai dari *evidence* dari jurusan RPL adalah 0,57 dan untuk jurusan TKJ yaitu 0,43.

- Perhitungan dinilai *evidence* pada setiap kriteria pada parameter tahun lulus sebagai berikut :

$$P_{(2021)} = \frac{\sum 2021}{\sum data} = \frac{506}{1533} = 0,33$$

$$P_{(2022)} = \frac{\sum 2022}{\sum data} = \frac{513}{1533} = 0,33$$

$$P_{(2023)} = \frac{\sum 2023}{\sum data} = \frac{514}{1533} = 0,34$$

Dari hasil perhitungan nilai *evidence* untuk kriteria tahun lulusan 2021 yaitu 0,33, untuk kriteria tahun lulusan 2022 yaitu 0,33 dan untuk kriteria tahun lulusan 2023 yaitu 0,34.

- Perhitungan dinilai *evidence* pada setiap kriteria pada parameter nilai UKK sebagai berikut :

$$P_{(UkkSK)} = \frac{\sum UKK \text{ Sangat Kompeten}}{\sum data} = \frac{539}{1533} = 0,35$$

$$P_{(UkkK)} = \frac{\sum UKK \text{ Kompeten}}{\sum data} = \frac{292}{1533} = 0,19$$

$$P_{(UkkCK)} = \frac{\sum UKK \text{ Cukup Kompeten}}{\sum data} = \frac{381}{1533} = 0,25$$

$$P_{(UkkTK)} = \frac{\Sigma Ukk \text{ Tidak Kompeten}}{\Sigma data} = \frac{321}{1533} = 0,21$$

Dari hasil nilai *evidence* untuk kriteria UKK dengan tingkat sangat kompeten mendapat nilai 0,35, untuk kriteria UKK dengan tingkat kompeten mendapat nilai 0,19, untuk kriteria UKK dengan tingkat cukup kompeten mendapat nilai 0,25, dan untuk kriteria UKK tidak kompeten mendapat nilai 0,21.

- Perhitungan dinilai *evidence* pada setiap kriteria pada parameter nilai KOMJARDAS sebagai berikut :

$$P_{(KomjardasA)} = \frac{\Sigma Komjardas A}{\Sigma data} = \frac{218}{1533} = 0,14$$

$$P_{(KomjardasB)} = \frac{\Sigma Komjardas B}{\Sigma data} = \frac{1192}{1533} = 0,78$$

$$P_{(KomjardasC)} = \frac{\Sigma Komjardas C}{\Sigma data} = \frac{123}{1533} = 0,08$$

$$P_{(KomjardasD)} = \frac{\Sigma Komjardas D}{\Sigma data} = \frac{0}{1533} = 0,00$$

$$P_{(KomjardasE)} = \frac{\Sigma Komjardas E}{\Sigma data} = \frac{0}{1533} = 0,00$$

Dari hasil nilai *evidence* untuk kriteria KOMJARDAS dengan tingkat A mendapat nilai 0,14, untuk kriteria KOMJARDAS dengan tingkat B mendapat nilai 0,78, untuk kriteria KOMJARDAS dengan tingkat C mendapat nilai 0,08, untuk kriteria KOMJARDAS dengan tingkat D mendapat nilai 0,00, dan untuk kriteria KOMJARDAS dengan tingkat E mendapat nilai 0,00.

- Perhitungan dinilai *evidence* pada setiap kriteria pada parameter nilai PROGDAS sebagai berikut :

$$P_{(ProgdasA)} = \frac{\Sigma Progdas A}{\Sigma data} = \frac{527}{1533} = 0,34$$

$$P_{(ProgdasB)} = \frac{\Sigma Progdas B}{\Sigma data} = \frac{908}{1533} = 0,60$$

$$P_{(ProgdasC)} = \frac{\Sigma Progdas C}{\Sigma data} = \frac{94}{1533} = 0,06$$

$$P_{(ProgdasD)} = \frac{\Sigma Progdas D}{\Sigma data} = \frac{2}{1533} = 0,001$$

$$P_{(ProgdasE)} = \frac{\Sigma Progdas E}{\Sigma data} = \frac{2}{1533} = 0,001$$

Dari hasil nilai *evidence* untuk kriteria PROGDAS dengan tingkat A mendapat nilai 0,34, untuk kriteria PROGDAS dengan tingkat B mendapat nilai 0,60, untuk kriteria PROGDAS dengan tingkat C mendapat nilai 0,06, untuk kriteria PROGDAS dengan tingkat D mendapat nilai 0,001, dan untuk kriteria PROGDAS dengan tingkat E mendapat nilai 0,001.

- Perhitungan dinilai *evidence* pada setiap kriteria pada parameter nilai SIMDIG sebagai berikut :

$$P_{(SimdigA)} = \frac{\Sigma Simdig A}{\Sigma data} = \frac{139}{1533} = 0,09$$

$$P_{(SimdigB)} = \frac{\Sigma Simdig B}{\Sigma data} = \frac{1320}{1533} = 0,86$$

$$P_{(SimdigC)} = \frac{\Sigma Simdig C}{\Sigma data} = \frac{72}{1533} = 0,05$$

$$P_{(SimdigD)} = \frac{\Sigma Simdig D}{\Sigma data} = \frac{0}{1533} = 0,00$$

$$P_{(SimdigE)} = \frac{\Sigma Simdig E}{\Sigma data} = \frac{2}{1533} = 0,001$$

Dari hasil nilai *evidence* untuk kriteria SIMDIG dengan tingkat A mendapat nilai 0,09, untuk kriteria SIMDIG dengan tingkat B mendapat nilai 0,86, untuk kriteria SIMDIG dengan tingkat C mendapat nilai 0,05, untuk kriteria SIMDIG dengan tingkat D mendapat nilai 0,00, dan untuk kriteria SIMDIG dengan tingkat E mendapat nilai 0,001.

➤ Perhitungan dinilai *evidence* pada setiap kriteria pada parameter nilai SISKOM sebagai berikut :

$$P_{(SiskomA)} = \frac{\sum Siskom A}{\sum data} = \frac{250}{1533} = 0,16$$

$$P_{(SiskomB)} = \frac{\sum Siskom B}{\sum data} = \frac{1112}{1533} = 0,72$$

$$P_{(SiskomC)} = \frac{\sum Siskom C}{\sum data} = \frac{87}{1533} = 0,056$$

$$P_{(SiskomD)} = \frac{\sum Siskom D}{\sum data} = \frac{1}{1533} = 0,0006$$

$$P_{(SiskomE)} = \frac{\sum Siskom E}{\sum data} = \frac{83}{1533} = 0,054$$

Dari hasil nilai *evidence* untuk kriteria SISKOM dengan tingkat A mendapat nilai 0,16, untuk kriteria SISKOM dengan tingkat B mendapat nilai 0,72, untuk kriteria SISKOM dengan tingkat C mendapat nilai 0,056, untuk kriteria SISKOM dengan tingkat D mendapat nilai 0,0006, dan untuk kriteria SISKOM dengan tingkat E mendapat nilai 0,054

#### 4.1.3 Perhitungan Nilai Likelihood

Perhitungan nilai *likelihood*  $P(x|c)$  didapatkan dari rumus sebagai berikut :

$$P(x_j | c_i) = \frac{\sum(x_1+x_2+\dots+x_j)}{\sum c_i} \quad (4.6)$$

Keterangan :

$$P(x_j | c_i) \quad = \text{nilai likelihood}$$

$$\sum(x_1 + x_2 + \dots + x_j) \quad = \text{Jumlah dari probabilitas fitur atau parameter linier}$$

$$\sum c_i \quad = \text{Jumlah dari probabilitas semua data}$$

Dengan dasar rumus *likelihood* tersebut, didapatkan dinilai *likelihood* pada setiap kriteria pada parameter sebagai berikut :

$$P_{(rpl|Linier)} = \frac{720}{1533} = 0,47$$

$$P_{(rpl|TidakLinier)} = \frac{152}{1533} = 0,099$$

$$P_{(tkj|Linier)} = \frac{545}{1533} = 0,36$$

$$P_{(tkj|TidakLinier)} = \frac{116}{1533} = 0,075$$

Dari hasil nilai *likelihood* untuk kriteria RPL yang linier mendapat nilai 0,47, untuk kriteria RPL yang tidak linier mendapat nilai 0,099, untuk kriteria TKJ yang linier mendapat nilai 0,36, dan untuk kriteria TKJ yang tidak linier mendapat nilai 0,075.

➤ Perhitungan dinilai *likelihood* pada setiap kriteria pada parameter tahun lulus sebagai berikut :

$$P_{(2021|Linier)} = \frac{406}{1533} = 0,26$$

$$P_{(2021|TidakLinier)} = \frac{100}{1533} = 0,065$$

$$P_{(2022|Linier)} = \frac{449}{1533} = 0,29$$

$$P_{(2022|TidakLinier)} = \frac{64}{1533} = 0,041$$

$$P_{(2023|Linier)} = \frac{410}{1533} = 0,27$$

$$P_{(2023|TidakLinier)} = \frac{104}{1533} = 0,067$$

Dari hasil nilai *likelihood* untuk kriteria tahun lulus 2021 yang linier mendapat nilai 0,26, untuk kriteria tahun lulus 2021 yang tidak linier mendapat nilai 0,065, untuk kriteria tahun lulus 2022 yang linier mendapat nilai 0,29, dan untuk kriteria tahun lulus 2022 yang tidak linier mendapat nilai 0,041, untuk kriteria tahun lulus 2023 yang linier mendapat nilai 0,27, dan untuk kriteria tahun lulus 2023 yang tidak linier mendapat nilai 0,067.

- Perhitungan dinilai *likelihood* pada setiap kriteria pada parameter nilai UKK sebagai berikut :

$$P_{(UkkSK|Linier)} = \frac{449}{1533} = 0,33$$

$$P_{(UkkSK|TidakLinier)} = \frac{90}{1533} = 0,058$$

$$P_{(UkkK|Linier)} = \frac{245}{1533} = 0,16$$

$$P_{(UkkK|TidakLinier)} = \frac{47}{1533} = 0,03$$

$$P_{(UkkCK|Linier)} = \frac{306}{1533} = 0,19$$

$$P_{(UkkCK|TidakLinier)} = \frac{75}{1533} = 0,048$$

$$P_{(UkkTK|Linier)} = \frac{265}{1533} = 0,17$$

$$P_{(UkkTK|TidakLinier)} = \frac{56}{1533} = 0,036$$

Dari hasil nilai *likelihood* untuk kriteria nilai UKK dengan predikat sangat kompeten yang linier mendapat nilai 0,34, untuk kriteria nilai UKK dengan predikat sangat kompeten yang tidak linier mendapat nilai 0,058, untuk kriteria nilai UKK dengan predikat kompeten yang linier mendapat nilai 0,16, dan untuk kriteria nilai UKK dengan predikat kompeten yang tidak linier mendapat nilai 0,03, untuk kriteria nilai UKK dengan predikat cukup kompeten yang linier mendapat nilai 0,19, dan untuk kriteria nilai UKK dengan cukup predikat kompeten yang tidak linier mendapat nilai 0,048, untuk kriteria nilai UKK dengan predikat tidak kompeten yang linier mendapat nilai 0,17, dan untuk kriteria nilai UKK dengan predikat tidak kompeten yang tidak linier mendapat nilai 0,036.

➤ Perhitungan dinilai *likelihood* pada setiap kriteria pada parameter nilai KOMJARDAS sebagai berikut :

$$P_{(KomjardasA|Linier)} = \frac{182}{1533} = 0,12$$

$$P_{(KomjardasA|TidakLinier)} = \frac{36}{1533} = 0,02$$

$$P_{(KomjardasB|Linier)} = \frac{976}{1533} = 0,64$$

$$P_{(KomjardasB|TidakLinier)} = \frac{216}{1533} = 0,14$$

$$P_{(KomjardasC|Linier)} = \frac{107}{1533} = 0,07$$

$$P_{(KomjardasC|TidakLinier)} = \frac{16}{1533} = 0,01$$

$$P_{(KomjardasD|Linier)} = \frac{0}{1533} = 0,00$$

$$P_{(KomjardasD|TidakLinier)} = \frac{0}{1533} = 0,00$$

$$P_{(KomjardasE|Linier)} = \frac{0}{1533} = 0,00$$

$$P_{(KomjardasE|TidakLinier)} = \frac{0}{1533} = 0,00$$

Dari hasil nilai *likelihood* untuk kriteria nilai KOMJARDAS dengan predikat A yang linier mendapat nilai 0,34, untuk kriteria nilai KOMJARDAS dengan predikat A yang tidak linier mendapat nilai 0,60, untuk kriteria nilai KOMJARDAS dengan predikat B yang linier mendapat nilai 0,06, dan untuk kriteria nilai KOMJARDAS dengan predikat B yang tidak linier mendapat nilai 0,001, untuk kriteria nilai KOMJARDAS dengan predikat C yang linier mendapat nilai 0,06, dan untuk kriteria nilai KOMJARDAS dengan predikat C yang tidak linier mendapat nilai 0,001, untuk kriteria nilai KOMJARDAS dengan predikat D yang linier mendapat nilai 0,06, dan untuk kriteria nilai KOMJARDAS dengan predikat D yang tidak

linier mendapat nilai 0,001, untuk kriteria nilai KOMJARDAS dengan predikat E yang linier mendapat nilai 0,06, dan untuk kriteria nilai KOMJARDAS dengan predikat E yang tidak linier mendapat nilai 0,001.

➤ Perhitungan dinilai *likelihood* pada setiap kriteria pada parameter nilai PROGDAS sebagai berikut :

$$P_{(ProgdasA|Linier)} = \frac{426}{1533} = 0,28$$

$$P_{(ProgdasA|TidakLinier)} = \frac{101}{1533} = 0,06$$

$$P_{(ProgdasB|Linier)} = \frac{757}{1533} = 0,49$$

$$P_{(ProgdasB|TidakLinier)} = \frac{151}{1533} = 0,09$$

$$P_{(ProgdasC|Linier)} = \frac{78}{1533} = 0,05$$

$$P_{(ProgdasC|TidakLinier)} = \frac{16}{1533} = 0,01$$

$$P_{(ProgdasD|Linier)} = \frac{2}{1533} = 0,001$$

$$P_{(ProgdasD|TidakLinier)} = \frac{0}{1533} = 0,00$$

$$P_{(ProgdasE|Linier)} = \frac{2}{1533} = 0,001$$

$$P_{(ProgdasE|TidakLinier)} = \frac{0}{1533} = 0,00$$

Dari hasil nilai *likelihood* untuk kriteria nilai PROGDAS dengan predikat A yang linier mendapat nilai 0,34, untuk kriteria nilai PROGDAS dengan predikat A yang tidak linier mendapat nilai 0,60, untuk kriteria nilai PROGDAS dengan predikat B yang linier mendapat nilai 0,06, dan untuk kriteria nilai PROGDAS dengan predikat B yang tidak linier mendapat nilai 0,001, untuk kriteria nilai PROGDAS dengan predikat C yang linier mendapat nilai 0,06, dan untuk kriteria nilai PROGDAS

dengan predikat C yang tidak linier mendapat nilai 0,001, untuk kriteria nilai PROGDAS dengan predikat D yang linier mendapat nilai 0,06, dan untuk kriteria nilai PROGDAS dengan predikat D yang tidak linier mendapat nilai 0,001, untuk kriteria nilai PROGDAS dengan predikat E yang linier mendapat nilai 0,06, dan untuk kriteria nilai PROGDAS dengan predikat E yang tidak linier mendapat nilai 0,001.

- Perhitungan dinilai *likelihood* pada setiap kriteria pada parameter nilai SIMDIG sebagai berikut :

$$P_{(SimdigA|Linier)} = \frac{110}{1533} = 0,06$$

$$P_{(SimdigA|TidakLinier)} = \frac{29}{1533} = 0,02$$

$$P_{(SimdigB|Linier)} = \frac{1097}{1533} = 0,71$$

$$P_{(SimdigB|TidakLinier)} = \frac{223}{1533} = 0,14$$

$$P_{(SimdigC|Linier)} = \frac{56}{1533} = 0,04$$

$$P_{(SimdigC|TidakLinier)} = \frac{16}{1533} = 0,01$$

$$P_{(SimdigD|Linier)} = \frac{0}{1533} = 0,00$$

$$P_{(SimdigD|TidakLinier)} = \frac{0}{1533} = 0,00$$

$$P_{(SimdigE|Linier)} = \frac{2}{1533} = 0,001$$

$$P_{(SimdigE|TidakLinier)} = \frac{0}{1533} = 0,00$$

Dari hasil nilai *likelihood* untuk kriteria nilai SIMDIG dengan predikat A yang linier mendapat nilai 0,34, untuk kriteria nilai SIMDIG dengan predikat A yang tidak linier mendapat nilai 0,60, untuk kriteria nilai SIMDIG dengan predikat B

yang linier mendapat nilai 0,06, dan untuk kriteria nilai SIMDIG dengan predikat B yang tidak linier mendapat nilai 0,001, untuk kriteria nilai SIMDIG dengan predikat C yang linier mendapat nilai 0,06, dan untuk kriteria nilai SIMDIG dengan predikat C yang tidak linier mendapat nilai 0,001, untuk kriteria nilai SIMDIG dengan predikat D yang linier mendapat nilai 0,06, dan untuk kriteria nilai SIMDIG dengan predikat D yang tidak linier mendapat nilai 0,001, untuk kriteria nilai SIMDIG dengan predikat E yang linier mendapat nilai 0,06, dan untuk kriteria nilai SIMDIG dengan predikat E yang tidak linier mendapat nilai 0,001.

➤ Perhitungan dinilai *likelihood* pada setiap kriteria pada parameter nilai SISKOM sebagai berikut :

$$P_{(SiskomA|Linier)} = \frac{203}{1533} = 0,13$$

$$P_{(SiskomA|TidakLinier)} = \frac{47}{1533} = 0,03$$

$$P_{(SiskomB|Linier)} = \frac{926}{1533} = 0,60$$

$$P_{(SiskomB|TidakLinier)} = \frac{186}{1533} = 0,12$$

$$P_{(SiskomC|Linier)} = \frac{65}{1533} = 0,04$$

$$P_{(SiskomC|TidakLinier)} = \frac{22}{1533} = 0,01$$

$$P_{(SiskomD|Linier)} = \frac{1}{1533} = 0,0006$$

$$P_{(SiskomD|TidakLinier)} = \frac{0}{1533} = 0,00$$

$$P_{(SiskomE|Linier)} = \frac{70}{1533} = 0,04$$

$$P_{(SiskomE|TidakLinier)} = \frac{13}{1533} = 0,008$$

Dari hasil nilai *likelihood* untuk kriteria nilai SISKOM dengan predikat A yang linier mendapat nilai 0,34, untuk kriteria nilai SISKOM dengan predikat A yang tidak linier mendapat nilai 0,60, untuk kriteria nilai SISKOM dengan predikat B yang linier mendapat nilai 0,06, dan untuk kriteria nilai SISKOM dengan predikat B yang tidak linier mendapat nilai 0,001, untuk kriteria nilai SISKOM dengan predikat C yang linier mendapat nilai 0,06, dan untuk kriteria nilai SISKOM dengan predikat C yang tidak linier mendapat nilai 0,001, untuk kriteria nilai SISKOM dengan predikat D yang linier mendapat nilai 0,06, dan untuk kriteria nilai SISKOM dengan predikat D yang tidak linier mendapat nilai 0,001, untuk kriteria nilai SISKOM dengan predikat E yang linier mendapat nilai 0,06, dan untuk kriteria nilai SISKOM dengan predikat E yang tidak linier mendapat nilai 0,001.

#### 4.1.4 Perhitungan Nilai *Posterior*

Perhitungan nilai posterior  $P(c|x)$  didapatkan dari rumus sebagai berikut :

$$Posterior = \frac{Likelihood \cdot Prior}{Evidance} \quad (4.7)$$

➤ Perhitungan dinilai *posterior* pada setiap kriteria pada parameter jurusan sebagai berikut :

$$P(linier|rpl) = \frac{P(rpl|linier) \cdot P(linier)}{P(rpl)} = \frac{0,47 \cdot 0,83}{0,57} = \frac{0,39}{0,57} = 0,68$$

$$P(linier|tkj) = \frac{P(tkj|linier) \cdot P(linier)}{P(tkj)} = \frac{0,36 \cdot 0,83}{0,43} = \frac{0,30}{0,43} = 0,69$$

Dari hasil nilai *posterior* untuk kriteria jurusan dengan tingkat RPL yang linier mendapat nilai 0,68, dan untuk kriteria jurusan dengan tingkat TKJ yang linier mendapat nilai 0,69.

- Perhitungan dinilai *posterior* pada setiap kriteria pada parameter tahun lulus sebagai berikut :

$$P(\text{linier}|2021) = \frac{P(2021|\text{linier}) * P(\text{linier})}{P(2021)} = \frac{0,26 * 0,83}{0,33} = \frac{0,215}{0,33} = 0,65$$

$$P(\text{linier}|2022) = \frac{P(2022|\text{linier}) * P(\text{linier})}{P(2022)} = \frac{0,29 * 0,83}{0,33} = \frac{0,24}{0,33} = 0,73$$

$$P(\text{linier}|2023) = \frac{P(2023|\text{linier}) * P(\text{linier})}{P(2023)} = \frac{0,27 * 0,83}{0,34} = \frac{0,22}{0,34} = 0,66$$

Dari hasil nilai *posterior* untuk kriteria tahun lulus dengan tingkat 2021 yang linier mendapat nilai 0,65, untuk kriteria tahun lulus dengan tingkat 2022 yang linier mendapat nilai 0,73, dan untuk kriteria tahun lulus dengan tingkat 2023 yang linier mendapat nilai 0,66.

- Perhitungan dinilai *posterior* pada setiap kriteria pada parameter nilai UKK sebagai berikut :

$$P(\text{linier}|UkkSK) = \frac{P(UkkSK|\text{linier}) * P(\text{linier})}{P(UkkSK)} = \frac{0,33 * 0,83}{0,35} = \frac{0,27}{0,35} = 0,78$$

$$P(\text{linier}|UkkK) = \frac{P(UkkK|\text{linier}) * P(\text{linier})}{P(UkkK)} = \frac{0,16 * 0,83}{0,19} = \frac{0,13}{0,19} = 0,70$$

$$P(\text{linier}|UkkCK) = \frac{P(UkkCK|\text{linier}) * P(\text{linier})}{P(UkkCK)} = \frac{0,19 * 0,83}{0,25} = \frac{0,16}{0,25} = 0,63$$

$$P(\text{linier}|UkkTK) = \frac{P(UkkTK|\text{linier}) * P(\text{linier})}{P(UkkTK)} = \frac{0,17 * 0,83}{0,21} = \frac{0,14}{0,21} = 0,67$$

Dari hasil nilai *posterior* untuk kriteria UKK dengan predikat sangat kompeten yang linier mendapat nilai 0,78, untuk kriteria UKK dengan tingkat kompeten yang linier mendapat nilai 0,70, untuk kriteria UKK dengan tingkat cukup kompeten yang linier mendapat nilai 0,63, dan untuk kriteria UKK dengan tingkat tidak kompeten yang linier mendapat nilai 0,67.

- Perhitungan dinilai *posterior* pada setiap kriteria pada parameter nilai KOMJARDAS sebagai berikut :

$$P(\text{linier}|\text{KomjardasA}) = \frac{P(\text{KomjardasA}|\text{linier}) * P(\text{linier})}{P(\text{KomjardasA})} = \frac{0,12 * 0,83}{0,14} = \frac{0,09}{0,14} = 0,71$$

$$P(\text{linier}|\text{KomjardasB}) = \frac{P(\text{KomjardasB}|\text{linier}) * P(\text{linier})}{P(\text{KomjardasB})} = \frac{0,64 * 0,83}{0,78} = \frac{0,53}{0,78} = 0,68$$

$$P(\text{linier}|\text{KomjardasC}) = \frac{P(\text{KomjardasC}|\text{linier}) * P(\text{linier})}{P(\text{KomjardasC})} = \frac{0,07 * 0,83}{0,08} = \frac{0,06}{0,08} = 0,72$$

$$P(\text{linier}|\text{KomjardasD}) = \frac{P(\text{KomjardasD}|\text{linier}) * P(\text{linier})}{P(\text{KomjardasD})} = \frac{0,00 * 0,83}{0,00} = \frac{0}{0,00} = 0$$

$$P(\text{linier}|\text{KomjardasE}) = \frac{P(\text{KomjardasE}|\text{linier}) * P(\text{linier})}{P(\text{KomjardasE})} = \frac{0,00 * 0,83}{0,00} = \frac{0}{0,00} = 0$$

Dari hasil nilai *posterior* untuk kriteria KOMJARDAS dengan predikat A yang linier mendapat nilai 0,71, untuk kriteria KOMJARDAS dengan predikat B yang linier mendapat nilai 0,68, untuk kriteria KOMJARDAS dengan predikat C yang linier mendapat nilai 0,72, untuk kriteria KOMJARDAS dengan predikat D yang linier mendapat nilai 0, dan untuk kriteria KOMJARDAS dengan predikat E yang linier mendapat nilai 0.

- Perhitungan dinilai *posterior* pada setiap kriteria pada parameter nilai PROGDAS sebagai berikut :

$$P(\text{linier}|\text{ProgdasA}) = \frac{P(\text{ProgdasA}|\text{linier}) * P(\text{linier})}{P(\text{ProgdasA})} = \frac{0,28 * 0,83}{0,34} = \frac{0,23}{0,34} = 0,68$$

$$P(\text{linier}|\text{ProgdasB}) = \frac{P(\text{ProgdasB}|\text{linier}) * P(\text{linier})}{P(\text{ProgdasB})} = \frac{0,49 * 0,83}{0,60} = \frac{0,41}{0,60} = 0,67$$

$$P(\text{linier}|\text{ProgdasC}) = \frac{P(\text{ProgdasC}|\text{linier}) * P(\text{linier})}{P(\text{ProgdasC})} = \frac{0,05 * 0,83}{0,06} = \frac{0,04}{0,06} = 0,69$$

$$P(\text{linier}|\text{ProgdasD}) = \frac{P(\text{ProgdasD}|\text{linier}) * P(\text{linier})}{P(\text{ProgdasD})} = \frac{0,001 * 0,83}{0,001} = \frac{0,00083}{0,001} = 0,83$$

$$P(\text{linier}|\text{ProgdasE}) = \frac{P(\text{ProgdasE}|\text{linier}) * P(\text{linier})}{P(\text{ProgdasE})} = \frac{0,001 * 0,83}{0,001} = \frac{0,00083}{0,001} = 0,83$$

Dari hasil nilai *posterior* untuk kriteria PROGDAS dengan predikat A yang linier mendapat nilai 0,68, untuk kriteria PROGDAS dengan predikat B yang linier mendapat nilai 0,67, untuk kriteria PROGDAS dengan predikat C yang linier mendapat nilai 0,69, untuk kriteria PROGDAS dengan predikat D yang linier mendapat nilai 0,83, dan untuk kriteria PROGDAS dengan predikat E yang linier mendapat nilai 0,83.

- Perhitungan dinilai *posterior* pada setiap kriteria pada parameter nilai SIMDIG sebagai berikut :

$$P(\text{linier}|\text{SimdigA}) = \frac{P(\text{SimdigA}|\text{linier}) * P(\text{linier})}{P(\text{SimdigA})} = \frac{0,06 * 0,83}{0,09} = \frac{0,05}{0,09} = 0,55$$

$$P(\text{linier}|\text{SimdigB}) = \frac{P(\text{SimdigB}|\text{linier}) * P(\text{linier})}{P(\text{SimdigB})} = \frac{0,71 * 0,83}{0,86} = \frac{0,58}{0,86} = 0,68$$

$$P(\text{linier}|\text{SimdigC}) = \frac{P(\text{SimdigC}|\text{linier}) * P(\text{linier})}{P(\text{SimdigC})} = \frac{0,04 * 0,83}{0,05} = \frac{0,03}{0,05} = 0,66$$

$$P(\text{linier}|\text{SimdigD}) = \frac{P(\text{SimdigD}|\text{linier}) * P(\text{linier})}{P(\text{SimdigD})} = \frac{0,00 * 0,83}{0,00} = \frac{0}{0,00} = 0$$

$$P(\text{linier}|\text{SimdigE}) = \frac{P(\text{SimdigE}|\text{linier}) * P(\text{linier})}{P(\text{SimdigE})} = \frac{0,001 * 0,83}{0,001} = \frac{0,00083}{0,001} = 0,83$$

Dari hasil nilai *posterior* untuk kriteria SIMDIG dengan predikat A yang linier mendapat nilai 0,55, untuk kriteria SIMDIG dengan predikat B yang linier mendapat nilai 0,68, untuk kriteria SIMDIG dengan predikat C yang linier mendapat nilai 0,66, untuk kriteria SIMDIG dengan predikat D yang linier mendapat nilai 0, dan untuk kriteria SIMDIG dengan predikat E yang linier mendapat nilai 0,83.

- Perhitungan dinilai *posterior* pada setiap kriteria pada parameter nilai SISKOM sebagai berikut :

$$P(\text{linier}|\text{SiskomA}) = \frac{P(\text{SiskomA}|\text{linier}) * P(\text{linier})}{P(\text{SiskomA})} = \frac{0,13 * 0,83}{0,16} = \frac{0,11}{0,16} = 0,67$$

$$P(\text{linier}|\text{SiskomB}) = \frac{P(\text{SiskomB}|\text{linier}) * P(\text{linier})}{P(\text{SiskomB})} = \frac{0,60 * 0,83}{0,72} = \frac{0,50}{0,72} = 0,69$$

$$P(\text{linier}|\text{SiskomC}) = \frac{P(\text{SiskomC}|\text{linier}) * P(\text{linier})}{P(\text{SiskomC})} = \frac{0,04 * 0,83}{0,056} = \frac{0,033}{0,056} = 0,59$$

$$P(\text{linier}|\text{SiskomD}) = \frac{P(\text{SiskomD}|\text{linier}) * P(\text{linier})}{P(\text{SiskomD})} = \frac{0,0006 * 0,83}{0,0006} = \frac{0,0005}{0,0006} = 0,83$$

$$P(\text{linier}|\text{SiskomE}) = \frac{P(\text{SiskomE}|\text{linier}) * P(\text{linier})}{P(\text{SiskomE})} = \frac{0,04 * 0,83}{0,054} = \frac{0,033}{0,054} = 0,61$$

Dari hasil nilai *posterior* untuk kriteria SISKOM dengan predikat A yang linier mendapat nilai 0,67, untuk kriteria SISKOM dengan predikat B yang linier mendapat nilai 0,69, untuk kriteria SISKOM dengan predikat C yang linier mendapat nilai 0,59, untuk kriteria SISKOM dengan predikat D yang linier mendapat nilai 0,83, dan untuk kriteria SISKOM dengan predikat E yang linier mendapat nilai 0,61.

## 4.2 Ujicoba

Pada langkah berikutnya, akan dilakukan uji coba penerapan algoritma *Naive Bayes* melalui strategi eksperimen. Dari hasil eksperimen tersebut, akan diperoleh performa algoritma *Naive Bayes* (akurasi, presisi, recall, dan f1-score).

Pengujian terhadap metode klasifikasi *Naive Bayes* dilakukan menggunakan teknik validasi silang dengan matriks kebingungan, di mana *dataset* yang digunakan dibagi menjadi dua bagian. Sebanyak 65% (996 baris) dari *dataset* digunakan sebagai data pelatihan, sementara 35% (537 baris) sisanya digunakan sebagai data uji.

Tabel 4.1 Data Testing Naive Bayes

No	Nama Siswa	Jurusan	Tahun Lulus	UKK	KOM JAR DAS	PROG DAS	SIM DIG	SIS KOM	Aktual	Prediksi
1	SYAUQI MUHAMMAD RAIS	TKJ 2	2022	TK	B	B	B	B	Tidak Linier	Linier
2	TITANIA AURELLIA PUTRI DWIENDRA	RPL 2	2023	K	B	B	B	B	Linier	Linier
3	ANGGIE VERAWATI ROHANASARI	TKJ 1	2023	CK	A	B	A	B	Linier	Linier
4	ANDHIKA BAGASKARA PUTRA NUSANTARA	RPL 5	2022	SK	B	B	B	A	Linier	Linier
5	ARIE FARCHAN FYRZATULLAH	TKJ 4	2023	SK	A	A	A	B	Linier	Linier
6	HANY SANINDA ZAHIRAL PUTRI	TKJ 4	2022	TK	B	B	B	A	Linier	Linier
7	ARFINA DEVI MAULIDAH	RPL 7	2022	SK	B	B	B	B	Linier	Linier
8	FATHIMAH AZZAHRA	TKJ 6	2021	K	B	A	B	A	Tidak Linier	Linier
9	BHISMA YUDHA SAPUTRA	RPL 5	2021	CK	B	A	B	E	Linier	Linier
10	Aradhea Hadwan	TKJ 1	2021	SK	B	B	B	B	Linier	Linier
11	MUHAMMAD BAGAS RAMADHAN	RPL 2	2021	K	B	A	B	E	Linier	Linier
12	AHMAD NUR FAHMI	RPL 5	2022	SK	B	B	B	B	Linier	Linier
13	NOVAL MISBAKHUS SURURI	TKJ 2	2023	K	B	B	A	B	Linier	Linier
14	SATRIYO YOGA PRADANA	TKJ 5	2021	K	A	A	B	B	Linier	Linier
15	ARSYA RAHMA AYULISTYA	RPL 4	2022	SK	B	A	B	B	Linier	Linier
16	CLEARASIA PRASTIKA SARI	RPL 2	2021	CK	B	B	B	B	Linier	Linier
17	Galur Arasy Lumintang	RPL 6	2021	CK	B	B	B	B	Linier	Linier
18	MUHAMMAD GENTA KURNIAWAN	RPL 1	2023	K	B	B	B	B	Linier	Linier
19	DEVIRA FANNY RYENA	RPL 2	2023	TK	C	B	B	B	Linier	Linier
20	FAIZA ANATHASYA EKA FALEN	RPL 7	2023	TK	C	C	B	B	Linier	Linier

No	Nama Siswa	Jurusan	Tahun Lulus	UKK	KOM JAR DAS	PROG DAS	SIM DIG	SIS KOM	Aktual	Prediksi
...	....	....	....	....	....	....	....	....	....	....
533	DEWI SARTIKA	TKJ 4	2021	TK	B	A	B	B	Linier	Linier
534	ACHMAD ZACKY GHOUTZU ZAMANI	TKJ 5	2023	SK	B	B	B	C	Linier	Linier
535	MUHAMMAD RAFFA ANDARRA PUTRA	RPL 5	2023	SK	B	B	B	B	Tidak Linier	Linier
536	FARAH NABILA ARRAHMAN	TKJ 3	2023	SK	B	B	C	B	Tidak Linier	Linier
537	DIMAS SUNU BAYU WIDODO	TKJ 1	2021	K	B	A	B	B	Tidak Linier	Linier

Setelah menerima hasil uji coba, langkah berikutnya adalah mengevaluasi kinerja metode klasifikasi *Naive Bayes* dengan menghitung tingkat akurasi. Dengan menggunakan matriks konfusi, kinerja metode *Naive Bayes* dapat dinilai. Evaluasi kinerja ini memanfaatkan matriks konfusi yang mengidentifikasi jumlah data uji yang diklasifikasikan dengan benar dan salah. Sebuah contoh matriks konfusi untuk klasifikasi biner dapat ditemukan dalam Tabel 4.2. tentang *confusion matrix*.

**Tabel 4.2 Confusion Matrix**

		Prediksi	
		Positif	Negatif
Data Aktual	Positif	TP	FP
	Negatif	FN	TN

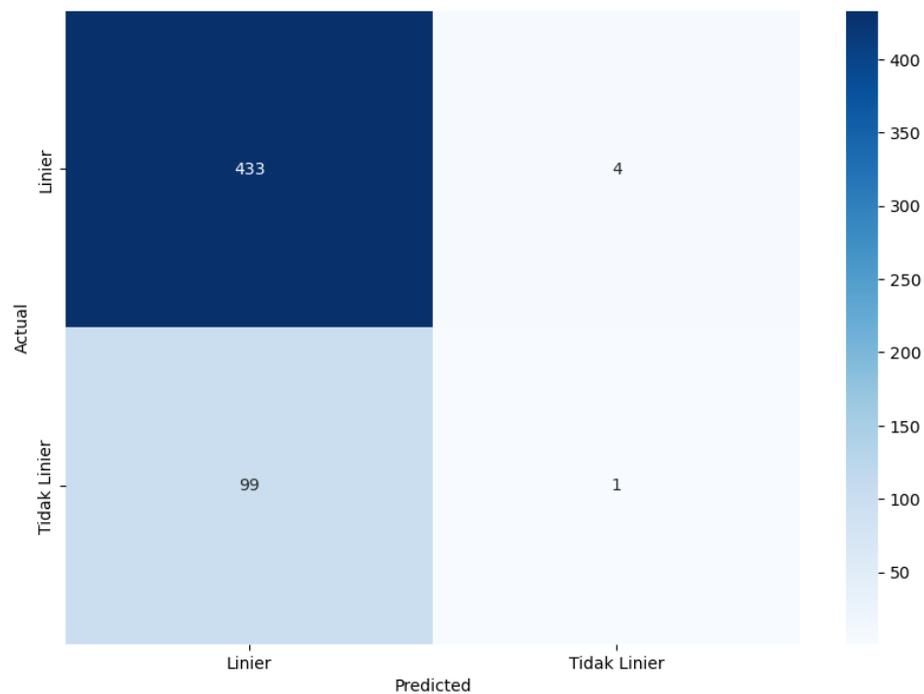
Keterangan:

TP (True Positive) adalah ketika sistem menghasilkan keputusan yang sesuai dengan hasil klasifikasi.

TN (True Negative) adalah ketika sistem tidak menghasilkan keputusan dan hasil klasifikasi menunjukkan kesesuaian.

FP (False Positive) adalah ketika sistem menghasilkan keputusan yang tidak ada dan hasil klasifikasi menunjukkan perbedaan.

FN (False Negative) adalah ketika sistem tidak menghasilkan keputusan namun hasil klasifikasi menunjukkan perbedaan.



**Gambar 4.1** *Confusion Matrix Naive Bayes*

Berdasarkan Gambar 4.1 tentang *Confusion Matrix Naive bayes* didapat nilai dari TP (*True Positive*) adalah 433, nilai FP (*False Positive*) adalah 4, nilai FN (*False Negative*) adalah 99, dan nilai TN (*True Negative*) adalah 1.

Rumus *confusion matrix* untuk menghitung *accuracy*, *precision*, *recall* dan *f1-score* seperti berikut :

$$Accuracy = \frac{TP+TN}{Total} \quad (4.7)$$

$$Precision = \frac{TP}{TP+FP} \quad (4.8)$$

$$Recall = \frac{TP}{TP+FN} \quad (4.9)$$

$$F1 - score = \frac{2TP}{2TP+FN+FP} \quad (4.10)$$

➤ Perhitungan hasil untuk mengukur performa dari metode *Naive Bayes* adalah :

$$Accuracy = \frac{TP+TN}{Total} = \frac{433+1}{537} = 0,80$$

$$Precision = \frac{TP}{TP+FP} = \frac{433}{433+4} = 0,99$$

$$Recall = \frac{TP}{TP+FN} = \frac{433}{433+99} = 0,81$$

$$F1 - score = \frac{2TP}{2TP+FN+FP} = \frac{2*433}{(2*433)+99+4} = \frac{866}{969} = 0,89$$

Berdasarkan hasil perhitungan tersebut, didapatkan nilai dari *accuracy* adalah 0,80, untuk nilai *precision* adalah 0,99, untuk nilai *recall* adalah 0,81, dan untuk nilai *f1-score* adalah 0,89.

### 4.3 Kesimpulan

Hasil analisa menggunakan algoritma *Naive Bayes*, dengan pendekatan eksperimental telah menghasilkan nilai akurasi, presisi, recall, dan f1-score seperti yang tercatat dalam Tabel 4.3 mengenai hasil eksperimen.

**Tabel 4.3 Hasil Eksperimen**

<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-score</i>
80 %	99 %	81 %	89 %

Berdasarkan Tabel 4.3 tentang hasil eksperimen didapatkan bahwa dengan menggunakan algoritma *Naive Bayes* menghasilkan nilai *accuracy* 80 % , nilai *precision* 99 %, nilai *recall* 81 %, dan nilai *f1-score* 89 %.

## BAB V

### METODE *DECISION TREE*

#### 5.1 Desain Metode

Metode *Decision Tree* (C4.5) merupakan salah satu teknik klasifikasi yang menggunakan pendekatan pohon keputusan untuk mengelompokkan data. Proses pembuatan pohon keputusan ini bergantung pada atribut yang signifikan dengan tingkat bobot yang tinggi. Penentuan node utama didasarkan pada perhitungan nilai entropy dan gain tertinggi sebagai pedoman dalam pembentukan pohon keputusan.

Dalam penelitian ini, metode Pohon Keputusan atau C4.5 yang digunakan dirancang sebagaimana yang tergambar pada Gambar 3.2 mengenai desain sistem. Terdapat 1533 baris data yang akan dianalisis. *Dataset* ini diperoleh dari Tabel 3.4 yang mencakup data alumni yang telah dibahas pada bab sebelumnya.

Secara umum, metode Pohon Keputusan (*Decision Tree*) C4.5 melibatkan serangkaian langkah dalam pembangunan struktur pohon keputusan, sebagai berikut:

- a. Menentukan atribut sebagai akar
- b. Membuat cabang untuk setiap nilai atribut
- c. Memisahkan kasus ke dalam cabang yang sesuai
- d. Mengulangi proses untuk setiap cabang hingga semua kasus dalam cabang tersebut memiliki kelas yang sama.

Untuk memilih atribut sebagai akar, didasarkan pada nilai *gain* tertinggi dari atribut-atribut yang ada. Nilai *Entropy* diperoleh dari persamaan berikut :

$$Entropy (S) = \sum_{i=1}^n - p_i * \log_2 p_i \quad (5.1)$$

Keterangan:

$S$  = Himpunan Kasus

$n$  = Jumlah partisi  $S$

$p_i$  = Proporsi dari  $S_i$  terhadap  $S$

➤ Untuk mendapatkan nilai *Gain* diperoleh dari persamaan berikut:

$$\text{Gain}(S, A) = \text{Entropy}(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * \text{Entropy}(S_i) \quad (5.2)$$

Keterangan :

$S$  = Himpunan Kasus

$A$  = Atribut

$n$  = Jumlah Partisi Atribut  $A$

$|S_i|$  = Jumlah kasus pada partisi ke- $i$

$|S|$  = Jumlah kasus dalam  $S$

Berikut perhitungan yang didapat dari rumus tersebut yang akan dituliskan pada tabel 5.1 tentang perhitungan *entropy* dan *gain*.

**Tabel 5.1 Perhitungan *Entropy* Dan *Gain***

<b>NODE</b>	<b>Atribut</b>	<b>Si</b>	<b>Jumlah Kasus (S)</b>	<b>Linier (Si)</b>	<b>Tidak Linier (Si)</b>	<b>Entropy</b>	<b>Gain</b>
1	Total		1533	1265	268	0,66861438	
	Jurusan	RPL	872	720	152	0,66747422	0,00070645
		TKJ	661	545	116	0,67011453	
	Tahun Lulus	2021	506	406	100	0,71715958	0,00687410
		2022	513	449	64	0,54287999	
		2023	514	410	104	0,72657285	
	Nilai UKK	Sangat Kompeten	539	449	90	0,65073989	0,00095238
		Kompeten	292	245	47	0,63659863	
		Cukup Kompeten	381	306	75	0,71558398	

		Tidak Kompeten	321	265	56	0,66779477	
Nilai KOMJAR DAS	A		218	182	36	0,6464599	0,00129093
	B		1192	976	216	0,682712697	
	C		123	107	16	0,557661111	
	D		0	0	0	0	
	E		0	0	0	0	
Nilai PROGDAS	A		527	426	101	0,704912173	0,00189041
	B		908	757	151	0,649169325	
	C		94	78	16	0,658191266	
	D		2	2	0	0	
	E		2	2	0	0	
Nilai SIMDIG	A		139	110	29	0,738861811	0,00211872
	B		1320	1097	223	0,655272801	
	C		72	56	16	0,764204507	
	D		0	0	0	0	
	E		2	2	0	0	
Nilai SISKOM	A		250	203	47	0,697268816	0,00291259
	B		1112	926	186	0,6514126	
	C		87	65	22	0,815800927	
	D		1	1	0	0	
	E		83	70	13	0,626177654	

## 5.2 Ujicoba

Langkah selanjutnya adalah menguji implementasi algoritma *Decision Tree* (C4.5) dengan menggunakan strategi pengujian. Hasil eksperimen digunakan untuk memperoleh nilai performa (*accuracy*, *precision*, *recall* dan *f1-score*) dari algoritma *Decision Tree*.

Pengujian metode klasifikasi *Decision Tree* (C4.5) dilakukan dengan teknik split validasi dengan *confusion matrix*, dimana pohon data yang digunakan dibagi menjadi dua bagian yaitu, diambil 65% (996 baris) dari kumpulan data akan

digunakan sebagai data pelatihan, diambil 35% (537 baris) dari kumpulan data digunakan sebagai data pengujian.

**Tabel 5.2 Data Testing Decision Tree**

No	Nama Siswa	Jurusan	Tahun Lulus	UKK	KOM JAR DAS	PROG DAS	SIM DIG	SIS KOM	Aktual	Prediksi
1	SYAUQI MUHAMMAD RAIS	TKJ 2	2022	TK	B	B	B	B	Tidak Linier	Linier
2	TITANIA AURELLIA PUTRI DWIENDRA	RPL 2	2023	K	B	B	B	B	Linier	Linier
3	ANGGIE VERAWATI ROHANASARI	TKJ 1	2023	CK	A	B	A	B	Linier	Linier
4	ANDHIKA BAGASKARA PUTRA NUSANTARA	RPL 5	2022	SK	B	B	B	A	Linier	Linier
5	ARIE FARCHAN FYRZATULLAH	TKJ 4	2023	SK	A	A	A	B	Linier	Linier
6	HANY SANINDA ZAHIRAL PUTRI	TKJ 4	2022	TK	B	B	B	A	Linier	Linier
7	ARFINA DEVI MAULIDAH	RPL 7	2022	SK	B	B	B	B	Linier	Linier
8	FATHIMAH AZZAHRA	TKJ 6	2021	K	B	A	B	A	Tidak Linier	Linier
9	BHISMA YUDHA SAPUTRA	RPL 5	2021	CK	B	A	B	E	Linier	Linier
10	Aradhea Hadwan	TKJ 1	2021	SK	B	B	B	B	Linier	Linier
11	MUHAMMAD BAGAS RAMADHAN	RPL 2	2021	K	B	A	B	E	Linier	Linier
12	AHMAD NUR FAHMI	RPL 5	2022	SK	B	B	B	B	Linier	Linier
13	NOVAL MISBAKHUS SURURI	TKJ 2	2023	K	B	B	A	B	Linier	Linier
14	SATRIYO YOGA PRADANA	TKJ 5	2021	K	A	A	B	B	Linier	Linier
15	ARSYA RAHMA AYULISTYA	RPL 4	2022	SK	B	A	B	B	Linier	Linier
16	CLEARASIA PRASTIKA SARI	RPL 2	2021	CK	B	B	B	B	Linier	Linier
17	Galur Arasy Lumintang	RPL 6	2021	CK	B	B	B	B	Linier	Linier
18	MUHAMMAD GENTA KURNIAWAN	RPL 1	2023	K	B	B	B	B	Linier	Linier

No	Nama Siswa	Jurusan	Tahun Lulus	UKK	KOM JAR DAS	PROG DAS	SIM DIG	SIS KOM	Aktual	Prediksi
19	DEVIRA FANNY RYENA	RPL 2	2023	TK	C	B	B	B	Linier	Linier
20	FAIZA ANATHASYA EKA FALEN	RPL 7	2023	TK	C	C	B	B	Linier	Tidak Linier
...	....	....	....	....	....	....	....	....	....	....
533	DEWI SARTIKA	TKJ 4	2021	TK	B	A	B	B	Linier	Linier
534	ACHMAD ZACKY GHOUTZU ZAMANI	TKJ 5	2023	SK	B	B	B	C	Linier	Tidak Linier
535	MUHAMMAD RAFFA ANDARRA PUTRA	RPL 5	2023	SK	B	B	B	B	Tidak Linier	Linier
536	FARAH NABILA ARRAHMAN	TKJ 3	2023	SK	B	B	C	B	Tidak Linier	Linier
537	DIMAS SUNU BAYU WIDODO	TKJ 1	2021	K	B	A	B	B	Tidak Linier	Linier

Berdasarkan *data testing* dari Tabel 5.2 tentang *data testing decision tree* didapatkan sebuah pohon keputusan sebagai berikut :

```

|--- Tahun Lulus_2022 > 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- SIMDIG_C <= 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- Jurusan_RPL 3 <= 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- Jurusan_TKJ 2 <= 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- SISKOM_B <= 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- PROGDAS_C <= 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- PROGDAS_B <= 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- Jurusan_RPL 4 <= 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- Jurusan_TKJ 4 <= 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- Jurusan_RPL 1 <= 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- Leaf
| samples = {'Linier': 0, 'Tidak Linier': 1}
|--- Jurusan_RPL 1 > 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}

```

```

|--- KOMJARDAS_B <= 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- Leaf
| samples = {'Linier': 0, 'Tidak Linier': 1}
|--- KOMJARDAS_B > 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- Leaf
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- Jurusan_TKJ 4 > 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- UKK_SK <= 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- Leaf
| samples = {'Linier': 1, 'Tidak Linier': 0}
|--- UKK_SK > 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- Leaf
| samples = {'Linier': 0, 'Tidak Linier': 1}
|--- Jurusan_RPL 4 > 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- KOMJARDAS_B <= 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- Leaf
| samples = {'Linier': 0, 'Tidak Linier': 1}
|--- KOMJARDAS_B > 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- Leaf
| samples = {'Linier': 1, 'Tidak Linier': 0}
|--- PROGDAS_B > 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- Leaf
| samples = {'Linier': 0, 'Tidak Linier': 1}
|--- PROGDAS_C > 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- Jurusan_TKJ 3 <= 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- Jurusan_TKJ 6 <= 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- Leaf
| samples = {'Linier': 0, 'Tidak Linier': 1}
|--- Jurusan_TKJ 6 > 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- Leaf
| samples = {'Linier': 1, 'Tidak Linier': 0}
|--- Jurusan_TKJ 3 > 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- Leaf
| samples = {'Linier': 1, 'Tidak Linier': 0}
|--- SISKOM_B > 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- KOMJARDAS_B <= 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- Jurusan_RPL 7 <= 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- UKK_TK <= 0.50

```

```

| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- Leaf
| samples = {'Linier': 0, 'Tidak Linier': 1}
|--- UKK_TK > 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- Leaf
| samples = {'Linier': 1, 'Tidak Linier': 0}
|--- Jurusan_RPL 7 > 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- UKK_K <= 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- PROGDAS_B <= 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- Leaf
| samples = {'Linier': 1, 'Tidak Linier': 0}
|--- PROGDAS_B > 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- Leaf
| samples = {'Linier': 0, 'Tidak Linier': 1}
|--- UKK_K > 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- Leaf
| samples = {'Linier': 1, 'Tidak Linier': 0}
|--- KOMJARDAS_B > 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- Jurusan_RPL 5 <= 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- PROGDAS_B <= 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- Jurusan_TKJ 6 <= 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- Jurusan_TKJ 3 <= 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- Jurusan_RPL 6 <= 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- Leaf
| samples = {'Linier': 0, 'Tidak Linier': 1}
|--- Jurusan_RPL 6 > 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- Leaf
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- Jurusan_TKJ 3 > 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- UKK_SK <= 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- Leaf
| samples = {'Linier': 0, 'Tidak Linier': 1}
|--- UKK_SK > 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- Leaf
| samples = {'Linier': 1, 'Tidak Linier': 0}
|--- Jurusan_TKJ 6 > 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- UKK_TK <= 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}

```

```

|--- Leaf
| samples = {'Linier': 0, 'Tidak Linier': 1}
|--- UKK_TK > 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- PROGDAS_C <= 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- Leaf
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- PROGDAS_C > 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- Leaf
| samples = {'Linier': 0, 'Tidak Linier': 1}
|--- PROGDAS_B > 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- Jurusan_TKJ 6 <= 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- UKK_TK <= 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- Jurusan_RPL 4 <= 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- Jurusan_RPL 7 <= 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- Jurusan_RPL 6 <= 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- UKK_CK <= 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- Leaf
| samples = {'Linier': 0, 'Tidak Linier': 1}
|--- UKK_CK > 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- Jurusan_RPL 2 <= 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- Leaf
| samples = {'Linier': 0, 'Tidak Linier': 1}
|--- Jurusan_RPL 2 > 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- Leaf
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- Jurusan_RPL 6 > 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- UKK_K <= 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- Leaf
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- UKK_K > 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- Leaf
| samples = {'Linier': 1, 'Tidak Linier': 0}
|--- Jurusan_RPL 7 > 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- UKK_K <= 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- Leaf
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- UKK_K > 0.50

```

```

        | samples = {'Linier': 0, 'Tidak Linier': 0}
        |--- Leaf
        | samples = {'Linier': 0, 'Tidak Linier': 1}
|--- Jurusan_RPL 4 > 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- Leaf
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- UKK_TK > 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- Jurusan_TKJ 5 <= 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- Jurusan_TKJ 4 <= 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- Jurusan_TKJ 3 <= 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- Leaf
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- Jurusan_TKJ 3 > 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- Leaf
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- Jurusan_TKJ 4 > 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- Leaf
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- Jurusan_TKJ 5 > 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- Leaf
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- Jurusan_TKJ 6 > 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- Leaf
| samples = {'Linier': 0, 'Tidak Linier': 1}
|--- Jurusan_RPL 5 > 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- UKK_SK <= 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- Leaf
| samples = {'Linier': 0, 'Tidak Linier': 1}
|--- UKK_SK > 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- PROGDAS_B <= 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- Leaf
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- PROGDAS_B > 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- Leaf
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- Jurusan_TKJ 2 > 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- UKK_CK <= 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- SISKOM_A <= 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}

```

```

|--- UKK_TK <= 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
| --- Leaf
| samples = {'Linier': 0, 'Tidak Linier': 1}
|--- UKK_TK > 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
| --- PROGDAS_C <= 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
| --- SISKOM_B <= 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
| --- Leaf
| samples = {'Linier': 1, 'Tidak Linier': 0}
|--- SISKOM_B > 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
| --- PROGDAS_A <= 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
| --- Leaf
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- PROGDAS_A > 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
| --- Leaf
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- PROGDAS_C > 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
| --- Leaf
| samples = {'Linier': 0, 'Tidak Linier': 1}
|--- SISKOM_A > 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
| --- Leaf
| samples = {'Linier': 0, 'Tidak Linier': 1}
|--- UKK_CK > 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
| --- SISKOM_B <= 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
| --- KOMJARDAS_B <= 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
| --- Leaf
| samples = {'Linier': 1, 'Tidak Linier': 0}
|--- KOMJARDAS_B > 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
| --- Leaf
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- SISKOM_B > 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
| --- Leaf
| samples = {'Linier': 0, 'Tidak Linier': 1}
|--- Jurusan_RPL 3 > 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
| --- Leaf
| samples = {'Linier': 0, 'Tidak Linier': 1}
|--- SIMDIG_C > 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
| --- Jurusan_RPL 2 <= 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
| --- SISKOM_B <= 0.50

```

```

| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- UKK_TK <= 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- Leaf
| samples = {'Linier': 1, 'Tidak Linier': 0}
|--- UKK_TK > 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- Leaf
| samples = {'Linier': 0, 'Tidak Linier': 1}
|--- SISKOM_B > 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- Leaf
| samples = {'Linier': 0, 'Tidak Linier': 1}
|--- Jurusan_RPL 2 > 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- UKK_K <= 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- Leaf
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- UKK_K > 0.50
| samples = {'Linier': 0, 'Tidak Linier': 0}
|--- Leaf
| samples = {'Linier': 1, 'Tidak Linier': 0}

```

Setelah hasil uji coba diperoleh, langkah berikutnya adalah mengevaluasi kinerja metode klasifikasi *Decision Tree* (C4.5) dengan menghitung tingkat akurasi. Penilaian kinerja metode ini dilakukan melalui *confusion matrix*, yang memberikan informasi mengenai keberhasilan metode *Decision Tree* (C4.5). *Confusion matrix* digunakan untuk mengevaluasi klasifikasi data uji yang benar dan yang salah. Sebagai contoh, *confusion matrix* untuk klasifikasi biner dapat dilihat pada Tabel 5.3 terkait *confusion matrix*.

**Tabel 5.3 Confusion Matrix**

		Prediksi	
		Positif	Negatif
Data Aktual	Positif	TP	FP
	Negatif	FN	TN

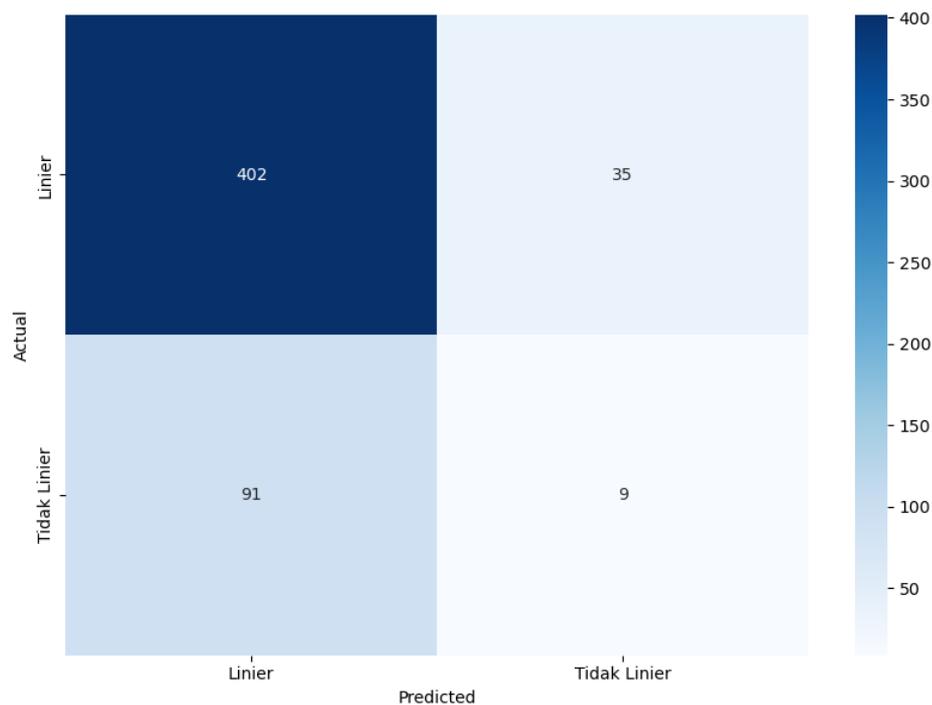
Keterangan:

TP (True Positive) adalah ketika sistem menghasilkan keputusan yang sesuai dengan hasil klasifikasi.

TN (True Negative) adalah ketika sistem tidak menghasilkan keputusan dan hasil klasifikasi menunjukkan kesesuaian.

FP (False Positive) adalah ketika sistem menghasilkan keputusan yang tidak ada dan hasil klasifikasi menunjukkan perbedaan.

FN (False Negative) adalah ketika sistem tidak menghasilkan keputusan namun hasil klasifikasi menunjukkan perbedaan



**Gambar 5.1** *Confusion Matrix Decision Tree*

Berdasarkan Gambar 5.1 tentang *Confusion Matrix Decision Tree* linearitas lulusan didapat nilai dari TP (*True Positive*) adalah 402, nilai FP (*False*

*Positive*) adalah 35, nilai FN (*False Negative*) adalah 91, dan nilai TN (*True Negative*) adalah 9.

Rumus *confusion matrix* untuk menghitung *accuracy*, *precision*, *recall* dan *f1-score* seperti berikut :

$$Accuracy = \frac{TP+TN}{Total} \quad (5.3)$$

$$Precision = \frac{TP}{TP+FP} \quad (5.4)$$

$$Recall = \frac{TP}{TP+FN} \quad (5.5)$$

$$F1 - score = \frac{2TP}{2TP+FN+FP} \quad (5.6)$$

➤ Hasil dari perhitungan performa metode *Decision Tree* sebagai berikut :

$$Accuracy = \frac{TP+TN}{Total} = \frac{402+9}{537} = 0,76$$

$$Precision = \frac{TP}{TP+FP} = \frac{402}{402+35} = 0,92$$

$$Recall = \frac{TP}{TP+FN} = \frac{402}{402+91} = 0,82$$

$$F1 - score = \frac{2TP}{2TP+FN+FP} = \frac{2 * 402}{(2 * 402) + 91 + 35} = \frac{804}{930} = 0,86$$

Berdasarkan hasil perhitungan tersebut, didapatkan nilai dari *accuracy* adalah 0,76, untuk nilai *precision* adalah 0,92, untuk nilai *recall* adalah 0,82, dan untuk nilai *f1-score* adalah 0,86.

### 5.3 Kesimpulan

Hasil analisa menggunakan algoritma *Decision Tree (C4.5)* dengan strategi eksperimen dapat dihasilkan nilai *accuracy*, *precision*, *recall* dan *f1-score* seperti terlihat pada Tabel 5.4. tentang hasil eksperimen

**Tabel 5.4 Hasil Eksperimen**

<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-score</i>
76 %	92 %	82 %	86 %

Berdasarkan Tabel 5.4 tentang hasil eksperimen didapatkan bahwa dengan menggunakan algoritma *Decision Tree (C4.5)* menghasilkan nilai *accuracy* 76 % , nilai *precision* 92 % , nilai *recall* 82 % , dan nilai *f1-score* 86 % .

## BAB VI PEMBAHASAN

### 6.1 Pembahasan Komparasi Performa Algoritma

Pada bagian ini, dilakukan perbandingan performa antara algoritma *Naive Bayes* dan algoritma *Decision Tree (C4.5)*. Evaluasi dilakukan berdasarkan nilai akurasi, presisi, recall, dan f1-score. Dari matriks kebingungan *Naive Bayes*, didapatkan nilai True Positive (TP) sebesar 433, False Positive (FP) sebesar 4, False Negative (FN) sebesar 99, dan True Negative (TN) sebesar 1. Sedangkan dari matriks kebingungan *Decision Tree*, ditemukan nilai TP sebesar 402, FP sebesar 35, FN sebesar 91, dan TN sebesar 9. Dengan data matriks kebingungan ini, dihitunglah nilai akurasi, presisi, recall, dan f1-score untuk setiap metode yang digunakan.

Berdasarkan diskusi tersebut, untuk mencapai kinerja optimal dari metode yang telah digunakan, disarankan untuk memilih yang memiliki tingkat ketepatan (*accuracy*) yang tertinggi. Perbandingan nilai *accuracy* dari kedua metode tersebut dapat dilihat secara *detail* dalam Tabel 6.1 mengenai perbandingan kinerja algoritma *Naive Bayes* dan *Decision Tree*.

**Tabel 6.1. Perbandingan performa algoritma *Naive Bayes* dengan *Decision Tree***

NO	Algoritma	Hasil Performa			
		<i>accuracy</i>	<i>precision</i>	<i>recall</i>	<i>f1-score</i>
1	<i>Naive Bayes</i>	80 %	99 %	81 %	89 %
2	<i>Decision Tree (C.45)</i>	76 %	92 %	82 %	86 %

Berdasarkan Tabel 6.1 mengenai perbandingan performa algoritma, disimpulkan bahwa pendekatan yang paling sesuai untuk melakukan klasifikasi

serapan lulusan adalah algoritma *Naive Bayes* dengan tingkat akurasi mencapai 80%, presisi sebesar 99%, recall sebesar 81%, dan f1-score sebesar 89%.

## 6.2 Manfaat klasifikasi linearitas serapan lulusan dalam pandangan agama

### Islam

Berdasarkan hasil pembahasan tersebut, didapatkan berbagai macam manfaat yang bisa dipetik untuk saling tolong menolong. Dengan adanya klasifikasi linearitas serapan lulusan, sekolah bisa melakukan kerjasama dengan industri dengan tepat sasaran. Selain itu, pihak sekolah juga bisa melakukan update kurikulum sesuai dengan kebutuhan industri. Tidak hanya itu, sekolah juga bisa memfasilitasi kebutuhan siswa dengan menyalurkan lulusan dari sekolah untuk bisa bekerja sesuai dengan kebutuhan dari alumninya. Allah *Subhanahu Wa Ta'ala* berfirman didalam Al-Qur'an surat Al Ma'idah/6:2

يَا أَيُّهَا الَّذِينَ آمَنُوا لَا تَحِلُّوا شَعَائِرَ اللَّهِ وَلَا الشَّهْرَ الْحَرَامَ وَلَا الْهُدْيَ وَلَا الْأَقْلَابَ وَلَا آمِنِينَ الْبَيْتِ  
الْحَرَامِ يَتَّبِعُونَ فَضْلًا مِّن رَّبِّهِمْ وَرِضْوَانًا وَإِذَا حَلَلْتُمْ فَاصْطَادُوا وَلَا يَجْرِمَنَّكُمْ شَنَا نُ قَوْمٍ أَن  
صَدُّوكُمْ عَنِ الْمَسْجِدِ الْحَرَامِ أَن تَعْتَدُوا وَتَعَاوَنُوا عَلَى الْبِرِّ وَالتَّقْوَىٰ وَلَا تَعَاوَنُوا عَلَى الْإِثْمِ  
وَالْعُدْوَانِ ۗ وَاتَّقُوا اللَّهَ إِنَّ اللَّهَ شَدِيدُ الْعِقَابِ

“Wahai orang-orang yang beriman, janganlah kamu melanggar syiar-syiar (kesucian) Allah, jangan (melanggar kehormatan) bulan-bulan haram, jangan (mengganggu) hadyu (hewan-hewan kurban) dan qalā'id (hewan-hewan kurban yang diberi tanda), dan jangan (pula mengganggu) para pengunjung Baitulharam sedangkan mereka mencari karunia dan rida Tuhannya! Apabila kamu telah bertahalul (menyelesaikan ihram), berburulah (jika mau). Janganlah sekali-kali kebencian(-mu) kepada suatu kaum, karena mereka menghalang-halangimu dari Masjidilharam, mendorongmu berbuat melampaui batas (kepada mereka). Tolong-menolonglah kamu dalam (mengerjakan) kebajikan dan takwa, dan jangan tolong-menolong dalam berbuat dosa dan permusuhan. Bertakwalah kepada Allah, sesungguhnya Allah sangat berat siksaan-Nya.” (QS. Al-Ma'idah [6:2])

Menurut tafsir Al-Qurtubi, disimpulkan bahwa ayat 2 surat Al-Maidah menegaskan pentingnya kerjasama dalam kebaikan dan ketakwaan, di mana semua makhluk diharapkan saling mendukung dan mendorong satu sama lain untuk melaksanakan perintah Allah *Subhanahu Wa Ta'ala*, serta menjauhi larangan-Nya. Dalam perspektif Islam, ayat ini menggarisbawahi pentingnya tolong-menolong dalam kebaikan sebagai ajaran yang telah diperintahkan oleh Allah *Subhanahu Wa Ta'ala*, seperti yang tertera dalam Al-Qur'an surat Al-Ma'idah ayat 2. Penelitian ini juga bertujuan membantu tim dari BKK sekolah dalam mencari industri yang sesuai dengan kebutuhan mereka.

## **BAB VII**

### **KESIMPULAN**

#### **7.1 Kesimpulan**

Dari kumpulan data alumni yang berjumlah 1533, digunakan sebagai *dataset* dengan berbagai parameter seperti Jurusan, Tahun Lulus, Bidang, Nilai Uji Kompetensi Kejuruan (UKK), Nilai Komputer Jaringan Dasar (KOMJARDAS), Nilai Pemrograman Dasar (PROGDAS), Nilai Simulasi Digital (SIMDIG), dan Nilai Sistem Komputer (SISKOM). *Dataset* ini dibagi menjadi dua bagian, yakni 996 baris untuk *data training* dan 537 baris untuk *data testing*. Selanjutnya, kedua *dataset* tersebut dianalisis menggunakan dua metode klasifikasi, yaitu *Naive Bayes* dan *Decision Tree*. Dari hasil analisis tersebut, dihasilkan tabel *confusion matrix* untuk setiap metode yang digunakan guna mengukur performa metode tersebut.

Berdasarkan hasil *confusion matrix*, evaluasi performa menggunakan algoritma *Naive Bayes* menunjukkan nilai akurasi sebesar 80%, presisi 81%, recall 99%, dan f1-score 89%. Sementara itu, evaluasi performa dengan menggunakan algoritma *Decision Tree* menghasilkan akurasi 76%, presisi 92%, recall 82%, dan f1-score 86%. Dari perbandingan tersebut, dapat disimpulkan bahwa algoritma *Naive Bayes* lebih tepat untuk mengklasifikasikan data lulusan berdasarkan performa klasifikasi, karena memiliki nilai akurasi yang lebih tinggi dibandingkan dengan algoritma *Decision Tree*.

## 7.2 Saran

Pengkajian lebih lanjut dapat dilakukan dengan menerapkan metode lain seperti *SVM*, *Fuzzy Model Tahani*, *Weighted Product*, *Neural Network*, dan *Profile Matching* menggunakan *dataset* yang sama. Hal ini bertujuan untuk membandingkan performa metode yang paling optimal.

## DAFTAR PUSTAKA

- Amalia, R. (2020). Penerapan Data Mining untuk Memprediksi Hasil Kelulusan Siswa Menggunakan Metode Naïve Bayes. *JUISI*, 06(01).
- Arief, S. (2022). *ULUMUL QUR'AN UNTUK PEMULA* (Vol. 1).
- CNN Indonesia. (2023, May 5). *Pengangguran di RI Terbanyak Lulusan SMK*. Retrieved 02/29/2024 from <https://www.cnnindonesia.com/ekonomi/20230505130917-92-945695/pengangguran-di-ri-terbanyak-lulusan-smk>
- Fakhrun Shiddieq, D., & Rizal Fadillah, M. (2019). *Penerapan Metode K-Means Untuk Klasifikasi Bidang Pekerjaan Alumni*. *JURNAL LPKIA* (Vol. 12).
- Farid Rifai, M., Jatnika, H., & Valentino, B. (2019). Penerapan Algoritma Naïve Bayes Pada Sistem Prediksi Tingkat Kelulusan Peserta Sertifikasi Microsoft Office Specialist (MOS), 12(2).
- Fortino, A. (2023). *Data Mining and Predictive Analytics for Business Decisions*.
- Hutapea, E., & Enggar Harususilo, Y. (2019, August 29). *SMK Bisa, Siswa SMK Indonesia Raih 15 Medali pada WSC 2019 Rusia*. Retrieved 02/29/2024 from <https://edukasi.kompas.com/read/2019/08/29/21175261/smk-bisa-siswa-smk-indonesia-raih-15-medali-pada-wsc-2019-rusia>
- IDN Times. (2023, October 20). *Potret Karya Siswa SMK di Penjuru Negeri*. Retrieved 02/29/2024 from <https://bali.idntimes.com/life/education/irma/potret-karya-siswa-smk-di-penjuru-negeri?page=all>
- Indra Purnama, D., Lathifah Islami, R., Sari, L., & Robinson Sihombing, P. (2021). Analisis Klasifikasi Data Tracer Study Dengan Support Vector Machine Dan Neural Network. *Jurnal Sistem Komputer Dan Kecerdasan Buatan*.
- Janu, S., Tyas, S., Febianah, M., Solikhah, F., Kamil, A. L., & Arifin, W. A. (2021). Analisis Perbandingan Algoritma Naive Bayes dan C.45 dalam Klasifikasi Data Mining untuk Memprediksi Kelulusan, 8(1).
- Kantardzic, M. (2020). *DATA MINING Concepts, Models, Methods, and Algorithms Third Edition*.
- Kartini, D., & Muliadi. (2016). Rancang Bangun Aplikasi K-Means untuk Klasifikasi Kelulusan Siswa Sekolah Kepolisian Negara Daerah Kalimantan Selatan.

- Kementerian Pendidikan dan Kebudayaan. (2018, February 1). *Terobosan Model Pembelajaran di SMK*. Retrieved 02/29/2024 from <https://www.kemdikbud.go.id/main/blog/2018/02/terobosan-model-pembelajaran-di-smk#:~:text=Tujuan%20pendidikan%20di%20SMK%20adalah,percepatan%20dan%20peningkatan%20kompetensi%20siswa.>
- Kotu, V., & Deshpande, B. (2015). *Predictive analytics and data mining : concepts and practice with RapidMiner*.
- Lukman, O. : (2018). *Kinerja Bursa Kerja Khusus (BKK) SMK Dalam Penyiapan Lulusannya Memasuki Dunia Kerja. Prodi Pendidikan Teknik Elektro* (Vol. 8). Retrieved from <http://journal.student.uny.ac.id/ojs/index.php/elektro>
- MALANG POSCO MEDIA. (2023, October 31). *Gandeng 35 Perusahaan, SMK PGRI 3 Malang Gelar Job Fair 2023; Fasilitasi Lulusan Masuk Dunia Kerja*. Retrieved 02/29/2024 from <https://malangposcomedia.id/gandeng-35-perusahaan-smk-pgri-3-malang-gelar-job-fair-2023-fasilitasi-lulusan-masuk-dunia-kerja/>
- Menteri Ketenagakerjaan Republik Indonesia. *Peraturan Menteri Ketenagakerjaan Republik Indonesia nomor 39 Tahun 2016 Tentang Penempatan Tenaga Kerja* (2016).
- Mukhlason, A., Winanti, T., & Yundra, E. (2020). *Analisa Indikator SMK Penyumbang Pengangguran di Provinsi Jawa Timur*.
- Prasojo, M., Sriyanto, & Triwidianti, J. (2021). *Prediksi Prestasi Siswa SMK Masuk Pasar Kerja Menggunakan Teknik Data Mining (Studi Kasus SMKN 1 Kota Agung Timur Tanggamus, Lampung)*.
- Presiden Republik Indonesia. *Undang-Undang Republik Indonesia nomor 20 Tahun 2003 Tentang Sistem Pendidikan Nasional* (2003). Available at: <https://peraturan.bpk.go.id/Details/43920/uu-no-20-tahun-2003>
- Ramadhan, L. A. (2022, June 11). *Siswa SMK Telkom Malang Disiapkan Jadi Pekerja Digital Profesional*. *Beritajatim.Com*. Retrieved 02/20/2024 from <https://beritajatim.com/siswa-smk-telkom-malang-disiapkan-jadi-pekerja-digital-profesional>
- Rizmayanti, A. I., Hidayati, N., Nugraha, F. S., Gata, W., & Mandiri, S. N. (2021). *Penerapan Data Mining Untuk Memprediksi Kompetensi Siswa Menggunakan Metode Decision Tree ( Studi Kasus SMK Multicomp Depok )*. *JURNAL SWABUMI*, 9(1), 2021.
- Said, A. (2023). *Berita Resmi Statistik*. Retrieved from <https://webapi.bps.go.id/download.php?f=2yo9/GXVDpdcQ2YjOsugqZU8/c>

zFtlw2hegol0umTtlh727beuilKARfDmzFhjeqlND6OJ9qMAd+6whUUIgk+  
XSyQJ3Bv6aLsWY+8kMxHRzWix6SHOJ4uLZKFFm8ux7tEkxDLoRByv6  
bjy+sqxd+zueNIWXJ+iFKreuvTbG7HyVd7j/4uP5sVTpgtOeoJmplpU0GCZ  
pGff4UcTG26GJy0pTUYv8LheX0l6CScqZ791KJ/8W5hUPe54vfLvm6rk  
+9EgsGSVnCVJ59peSNuEjX2UnDr8rW/2LIFpzJG0gDUI=

Shihab, M. Q. (2002). Tafsir Al-Mishbah : pesan, kesan dan keserasian Al-Qur'an / M. Quraish Shihab (Vol. 15).

Sinaga, K., & Handoko, K. (2021). *Implementasi Data Mining Untuk Memprediksi Kelulusan Siswa Dengan Metode Naive Bayes*. *JURNAL COMASIE* (Vol. 04).

Suhandi, Wiguna, W., & Quraysin, I. (2021). Dinamika Permasalahan Ketenagakerjaan dan Pengangguran di Indonesia. <https://doi.org/10.46306/vls.v1i1>

Ula, M., Zulhusna, R., Putra Fhonna, R., & Pratama, A. (2022). Penerapan Model Klasifikasi K-Nearest Neighbor Dalam Pencarian Kesesuaian Pekerjaan. *METIK JURNAL*, 6(1), 18–23. <https://doi.org/10.47002/metik.v6i1.343>

Widyawati, S. (2023, October 4). *SMKN 11 Kota Malang Gelar Job Fair, Diikuti 22 Industri Kreatif & Umum, Serap Lulusan SMK Siap Kerja*. Retrieved 02/29/2024 from <https://jatim.tribunnews.com/2023/10/04/smkn-11-kota-malang-gelar-job-fair-diikuti-22-industri-kreatif-umum-serap-lulusan-smk-siap-kerja>

Wiwik Sri Rahayu Ginantra, N. L., Nur Arifah, F., Hadi Wijaya, A., Septarini, R. S., Ahmad Dewa Putu Yudhi Ardiana, N., Effendy, F., ... Surya Negara, E. (2021). Data Mining dan Penerapan Algoritma.

Wydiastuty Kusuma, L. (2019). *Prediksi Kemampuan Lulusan SMK untuk Dapat Bersaing Di Dunia Kerja dengan Menggunakan Naïve Bayes: Studi Kasus SMK Buddhi Tangerang*. *JURNAL ALGOR* (Vol. 1). Retrieved from <https://jurnal.buddhidharma.ac.id/index.php/algor/index>