

**KLASIFIKASI BIDANG KEAHLIAN MAHASISWA
MENGUNAKAN ALGORITMA *NAIVE BAYES***

SKRIPSI

Program Studi Sarjana Informatika
Jurusan Informatika

Oleh:

FAJAR ISTIGHFAR

NIM D1041161002



FAKULTAS TEKNIK
UNIVERSITAS TANJUNGPURA
PONTIANAK
2022

HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini:

Nama : Fajar Istighfar

NIM : D1041161002

menyatakan bahwa dalam skripsi yang berjudul “KLASIFIKASI BIDANG KEAHLIAN MENGGUNAKAN ALGORITMA *NAIVE BAYES*” tidak terdapat karya yang pernah diajukan untuk memperoleh gelar sarjana di suatu perguruan tinggi manapun. Sepanjang pengetahuan Saya, tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam Daftar Pustaka.

Demikian pernyataan ini dibuat dengan sebenar-benarnya. Saya sanggup menerima konsekuensi akademis dan hukum di kemudian hari apabila pernyataan yang dibuat ini tidak benar.

Pontianak, 31 Januari 2022

Fajar Istighfar
NIM D1041161002



KEMENTERIAN PENDIDIKAN, KEBUDAYAAN,
RISET, DAN TEKNOLOGI
UNIVERSITAS TANJUNGPURA
FAKULTAS TEKNIK

Jl. Prof. Dr. H. Hadari Nawawi Pontianak 78124
Telepon (0561) 740186 Email : ft@untan.ac.id Website : teknik.untan.ac.id

HALAMAN PENGESAHAN

**KLASIFIKASI BIDANG KEAHLIAN MAHASISWA
MENGUNAKAN ALGORITMA *NAIVE BAYES***

Program Studi Sarjana Informatika
Jurusan Informatika

Oleh:

Fajar Istighfar
NIM D1041161002

Telah dipertahankan di depan Penguji Skripsi pada tanggal 31 Januari 2022 dalam sidang secara daring (*Online*) dan diterima sebagai salah satu persyaratan untuk memperoleh gelar sarjana.

Susunan Penguji Skripsi:

Dosen Pembimbing Utama	: Dr. Arif Bijaksana PN, S.T., M.T.	(NIP. 197208081998021002)
Dosen Pembimbing Kedua	: Tursina, S.T., M.Cs.	(NIP. 197801152002122003)
Dosen Penguji Utama	: Dr. Herry Sujaini, S.T., M.T.	(NIP. 196806291997021001)
Dosen Penguji Kedua	: Rina Septiriana, S.T., M.Cs.	(NIP. 198709232020122001)

Pontianak, 31 Januari 2022
Dekan,

Wakil Dekan Bidang Akademik

Dr. rer.nat. Ir. R. M. Rustamaji, M.T., IPU
NIP 196801161994031003

Dr.-Ing. Ir. Slamet Widodo M.T., IPM
NIP. 196712231992031002

HALAMAN PERSEMBAHAN

Skripsi ini didedikasikan kepada kedua orang tua saya (Bapak Endang Safaat dan Ibu Ramlah) dan saudariku Pertiwi Lara Istighfar, serta seluruh kawan kawan Angkatan 2016 Fakultas Teknik Universitas Tanjungpura. Apa yang telah mereka berikan melebihi daripada apa yang pernah saya inginkan.

KATA PENGANTAR

Puji dan syukur kepada Tuhan Yang Maha Esa, atas keberkahan-Nya saya dapat menyelesaikan penulisan skripsi ini. Laporan penelitian ini disusun sebagai syarat untuk menyelesaikan jenjang pendidikan Strata 1 (S1) pada Fakultas Teknik, Program Studi Informatika, Universitas Tanjungpura. Judul yang diajukan adalah “Klasifikasi Bidang Keahlian Mahasiswa Menggunakan Algoritma *Naive Bayes*”.

Melalui laporan ini, ucapan terima kasih saya sampaikan kepada pihak-pihak sebagai berikut:

1. Bapak Dr. Arif Bijaksana Putra Negara, S.T., M.T. selaku pembimbing akademik sekaligus pembimbing utama.
2. Ibu Tursina, S.T., M.Cs. selaku pembimbing pendamping.
3. Bapak Dr. Herry Sujaini, S.T., M.T. selaku penguji utama.
4. Ibu Rina Septiriana, S.T., M.Cs. selaku penguji pendamping.
5. Bapak M. Azhar Irwansyah, S.T., M.Eng. selaku Ketua Program Studi Informatika Fakultas Teknik Universitas Tanjungpura.
6. Bapak Dr. Yus Sholva, S.T., M.T. selaku Ketua Jurusan Informatika Fakultas Teknik Universitas Tanjungpura.
7. Bapak Dr. rer.nat. Ir. R. M. Rustamaji, M.T., IPU selaku Dekan Fakultas Teknik Universitas Tanjungpura.

Penulis menyadari banyak terdapat kekurangan dalam skripsi ini, oleh karena itu penulis mengharapkan kritik dan saran yang membangun. Semoga skripsi ini dapat bermanfaat bagi semua pihak, terima kasih.

Pontianak, 22 Januari 2022
Penulis,

Fajar Istighfar

ABSTRAK

Salah satu faktor yang dapat menentukan pilihan bidang keahlian mahasiswa yang dimiliki Jurusan Informatika Fakultas Teknik Universitas Tanjungpura adalah penguasaan sub kemampuan yang dikuasai oleh mahasiswa terhadap kriteria sub kemampuan pada bidang pilihan tersebut. Apabila mahasiswa memiliki kemampuan yang dikuasai di bidang keahliannya tentunya akan mempermudah mahasiswa dalam proses pemilihan topik tugas akhir/skripsi. Penelitian ini bertujuan untuk menerapkan dan menguji performa klasifikasi algoritma *Naive Bayes* dalam mengklasifikasikan bidang keahlian untuk mendapatkan model klasifikasi bidang keahlian mahasiswa berdasarkan kemampuan yang dikuasai oleh mahasiswa terhadap kriteria sub kemampuan pada kelompok keahlian Jurusan Informatika Fakultas Teknik Universitas Tanjungpura. Metode yang digunakan dalam penelitian ini adalah *Synthetic Minority Oversampling Technique* (SMOTE) sebagai *preprocessing* untuk menangani ketidakseimbangan kelas, *Chi Square* sebagai *feature selection* dalam mengurangi dimensi data, menghilangkan fitur-fitur yang tidak relevan dan *Naive Bayes* yang berfungsi untuk mengklasifikasi data. Pengujian akurasi metode pada penelitian ini dilakukan dengan *K-Fold Cross Validation* dan evaluasi menggunakan *Confusion Matrix*. Pengujian dilakukan dengan menggunakan empat skenario pengujian. Skenario pengujian pertama dilakukan menggunakan algoritma *Naive Bayes* dengan hasil nilai akurasi sebesar 66%. Skenario pengujian kedua dilakukan menggunakan algoritma *Naive Bayes* dan *Chi Square* dengan hasil nilai akurasi sebesar 71%. Skenario pengujian ketiga dilakukan menggunakan algoritma *Naive Bayes* dan *SMOTE* dengan hasil nilai akurasi sebesar 79%. Skenario pengujian keempat dilakukan menggunakan algoritma *Naive Bayes*, *SMOTE* dan *Chi Square* dengan hasil nilai akurasi sebesar 80%. Terdapat peningkatan akurasi pada klasifikasi dengan penambahan *SMOTE* dan *Chi Square* sebagai *balanced data* dan *feature selection*. Hal ini dikarenakan data set yang digunakan sudah seimbang dan memiliki kriteria yang relevan, dari 228 data set sebanyak 183 data berhasil diprediksi dengan benar. Sehingga model dengan menggunakan algoritma *Naive Bayes*, *SMOTE* dan *Chi Square* dapat memberikan performa model yang baik dan dapat digunakan dalam memprediksi bidang keahlian mahasiswa.

Kata kunci: Bidang Keahlian, klasifikasi *Naive Bayes*, *SMOTE*, *Chi Square*, *K-Fold Cross Validation*, *Confusion Matrix*

ABSTRACT

One of the factors that can determine the choice of the student's field of expertise owned by the Department of Informatics, Faculty of Engineering, Tanjungpura University is the mastery of the sub-ability controlled by the student against the criteria of sub-ability in the chosen field. Suppose students can master in their field of expertise. In that case, it will undoubtedly make it easier for students to select the final project/thesis topic. This study aims to apply and test the classification performance of the Naive Bayes algorithm in classifying areas of expertise to obtain a classification model for students' areas of expertise based on the abilities mastered by students against the sub-ability criteria in the expertise group of the Department of Informatics, Faculty of Engineering, Tanjungpura University. The method used in this research is Synthetic Minority Oversampling Technique (SMOTE) as preprocessing to deal with class imbalance, Chi-Square as feature selection in reducing data dimensions, eliminating irrelevant features, and Naive Bayes, which functions to classify data. Testing the method's accuracy in this study was conducted using K-Fold Cross-Validation and evaluation utilizing the Confusion Matrix. The test is carried out using four test scenarios. The first test scenario was carried out using the Naive Bayes algorithm with an accuracy value of 66%. The second test scenario was carried out using the Naive Bayes and Chi-Square algorithms with an accuracy value of 71%. The third test scenario was carried out using the Naive Bayes algorithm and SMOTE with an accuracy value of 79%. The fourth test scenario uses the Naive Bayes, SMOTE, and Chi-Square algorithms with an accuracy value of 80%. There is an increase in classification accuracy with SMOTE and Chi-Square as balanced data and feature selection. This is because the data set used is balanced and has relevant criteria; from 228 data sets, the model successfully predicted 183 data correctly so that the model using the Naive Bayes, SMOTE and Chi Square algorithms can provide a good model performance and can be used to predict students' areas of expertise.

Keywords: Field of Expertise, Naive Bayes Classification, SMOTE, Chi Square, K-fold cross validation, Confusion Matrix

DAFTAR ISI

Halaman Pernyataan	ii
Halaman Pengesahan.....	iii
Halaman Persembahan.....	iv
Kata Pengantar.....	v
Abstrak.....	vi
Abstract.....	vii
Daftar Isi	viii
Daftar Tabel.....	x
Daftar Gambar	xi
Daftar Kode Program	xii
Daftar Lampiran	xiii
Bab I Pendahuluan.....	1
1.1 Latar Belakang.....	1
1.2 Perumusan Masalah.....	4
1.3 Tujuan Penelitian.....	4
1.4 Pembatasan Masalah	4
1.5 Sistematika Penulisan	5
Bab II Tinjauan Pustaka	6
2.1 Kajian Terkait	6
2.2 Jurusan Informatika Fakultas Teknik Universitas Tanjungpura	9
2.3 <i>Machine Learning</i>	10
2.4 Klasifikasi.....	12
2.5 Bahasa Pemrograman Python.....	13
2.6 Jupyter Notebook.....	14
2.7 <i>Scikit-Learn</i>	14
2.8 Algoritma <i>Naive Bayes</i>	14
2.9 <i>Bernoulli Naive Bayes</i>	15
2.10 <i>Synthetic Minority Oversampling Technique (SMOTE)</i>	16
2.11 <i>Chi Square</i>	16
2.12 <i>K-Fold Cross Validation</i>	17
2.13 Pengujian <i>Confusion Matrix</i>	17
Bab III Metodologi Penelitian	19
3.1 Bahan Penelitian.....	19
3.2 Perangkat Pendukung Penelitian	19
3.2.1 Perangkat Lunak.....	19
3.2.2 Perangkat Keras.....	20
3.3 Metode Penelitian	20

3.3.1	Analisis Kebutuhan	23
3.3.2	Pengumpulan Data	26
3.3.3	Pengolahan Data.....	28
3.3.3.1	Pelabelan.....	28
3.3.3.2	Inisialisasi	30
3.3.4	<i>Preprocessing</i>	31
3.3.4.1	<i>Cleaning</i>	31
3.3.4.2	<i>Ordinal Encoding</i>	32
3.3.4.3	<i>Data Selection</i>	33
3.3.4.4	<i>SMOTE</i>	34
3.3.4.5	<i>Feature Selection</i>	36
3.3.5	Pemodelan	38
3.3.6	<i>Validation</i>	39
3.3.7	Pengujian.....	40
Bab IV	Hasil dan Analisis.....	42
4.1	Hasil Pengujian.....	42
4.2	Analisis Hasil Pengujian.....	47
Bab V	Kesimpulan dan Saran	49
5.1	Kesimpulan.....	49
5.2	Saran	50
Daftar Pustaka.....		51

DAFTAR TABEL

Tabel 2. 1 Tabel Perbandingan Penelitian.....	8
Tabel 2. 2 <i>Confusion Matrix</i>	18
Tabel 3. 1 Kriteria Sub Kemampuan.....	24
Tabel 3. 2 Data Hasil Survei Kemampuan	27
Tabel 3. 3 Pelabelan	29
Tabel 3. 4 Skenario Pengujian.....	40
Tabel 4. 1 Hasil Pengujian.....	46

DAFTAR GAMBAR

Gambar 2. 1 Cara kerja <i>machine learning</i>	11
Gambar 3. 1 Metode penelitian	20
Gambar 3. 2 Hasil inisialisasi	30
Gambar 3. 3 Hasil <i>cleaning</i>	32
Gambar 3. 4 Hasil <i>ordinal encoding</i>	33
Gambar 3. 5 Hasil <i>data selection</i>	34
Gambar 3. 6 Data sebelum <i>smote</i>	35
Gambar 3. 7 Data sesudah <i>smote</i>	35
Gambar 3. 8 Hasil <i>smote</i>	36
Gambar 3. 9 Hasil <i>chi square</i>	38
Gambar 3. 10 Tahapan <i>cross validation</i>	40
Gambar 4. 1 Hasil skenario pengujian 1.....	42
Gambar 4. 2 Hasil skenario pengujian 2.....	43
Gambar 4. 3 Hasil skenario pengujian 3.....	44
Gambar 4. 4 Hasil skenario pengujian 4.....	45

DAFTAR KODE PROGRAM

Kode program 3. 1 Inisialisasi	30
Kode program 3. 2 Menyimpan Hasil <i>Preprocessing</i>	31
Kode program 3. 3 <i>Cleaning</i>	31
Kode program 3. 4 <i>Ordinal Encoding</i>	33
Kode program 3. 5 <i>Data Selection</i>	34
Kode program 3. 6 <i>SMOTE</i>	35
Kode program 3. 7 Menyimpan hasil <i>smote</i>	36
Kode program 3. 8 <i>Chi Square</i>	37
Kode program 3. 9 <i>BernoulliNB</i>	39
Kode program 3. 10 <i>K-Fold</i>	40
Kode program 3. 11 <i>Confusion Matrix</i>	41

DAFTAR LAMPIRAN

Lampiran A	KUESIONER KEMAMPUAN INFORMATIKA	A-1
Lampiran B	KUESIONER KEMAMPUAN MAHASISWA.....	B-1
Lampiran C	HASIL KUESIONER KEMAMPUAN INFORMATIKA	C-1
Lampiran D	HASIL KUESIONER KEMAMPUAN MAHASISWA.....	D-1
Lampiran E	PELABELAN.....	E-1
Lampiran F	HASIL <i>CONFUSION MATRIX</i>	F-1

BAB I

PENDAHULUAN

1.1 Latar Belakang

Jurusan Informatika Fakultas Teknik Universitas Tanjungpura hingga saat ini masih menggunakan Kurikulum Kerangka Kualifikasi Nasional Indonesia (KKNI) tahun 2016. Dalam mempersiapkan lulusannya untuk terjun di masyarakat sesuai dengan standar kompetensi kerja baik nasional maupun internasional, khususnya di bidang Informatika. Jurusan Informatika mensyaratkan pembelajaran baik teori untuk penguasaan pengetahuan, praktik untuk penguasaan kemampuan kerja atau *skills*, kecakapan hidup dan kemampuan manajerial. Setiap mahasiswa akan melalui beberapa proses dalam perkuliahan untuk mendapatkan gelar sarjana, tugas akhir atau skripsi adalah bagian dari proses perkuliahan tersebut. Pada Jurusan Informatika, tugas akhir ini memerlukan tahapan dalam penyelesaiannya diantaranya adalah pengajuan proposal tugas akhir.

Skripsi merupakan karya akhir atau karya puncak yang dianggap bisa memberikan indikator kadar pemahaman atau ketercapaian disiplin ilmu mahasiswa yang bersangkutan. Pada tingkat perguruan tinggi seseorang belajar sesuai bidang keahlian tertentu, sehingga *output* dari sebuah perguruan tinggi diharapkan mampu menghasilkan sumber daya manusia yang berkualitas sesuai dengan bidang keahlian mahasiswa berdasarkan kemampuan yang dikuasai. Oleh sebab itu pengelompokan bidang keahlian mahasiswa berdasarkan kemampuan yang dikuasai sangat dibutuhkan (Munslich, 2008).

Bidang keahlian merupakan bagian dari kurikulum berbasis kompetensi. Bidang keahlian adalah kumpulan dari beberapa mata kuliah pendukung yang akan mengantarkan mahasiswa menuju proses penyelesaian skripsi (Fuad, 2019). Adapun kelompok keahlian yang dimiliki Jurusan Informatika Fakultas Teknik Universitas Tanjungpura yaitu:

1. Sistem Informasi dan Data Spasial
2. Rekayasa Perangkat Lunak dan Komputasi Bergerak
3. Jaringan dan Keamanan
4. Komputasi dan Kecerdasan Buatan

Terdapat beberapa faktor yang dapat menentukan bidang keahlian mahasiswa, seperti nilai akademik, minat mahasiswa, dan kemampuan (*skill*) mahasiswa itu sendiri. Mengingat pentingnya bidang keahlian bagi mahasiswa, sebuah solusi yang dapat membantu meminimalisir dampak dari kekeliruan seperti kesulitan memahami permasalahan, kesulitan memecahkan persoalan, dan ketidakmampuan untuk mandiri dalam proses menyelesaikan skripsi. Untuk itu diperlukan suatu metode klasifikasi sebagai sebuah metode yang mampu mengelompokkan bidang keahlian mahasiswa berdasarkan kemampuan yang dikuasai ke dalam kelompok keahlian. Klasifikasi merupakan salah satu metode yang sering digunakan di data *mining*. Klasifikasi termasuk ke dalam *supervised learning* karena dalam proses klasifikasi terdapat proses pembelajaran dengan data yang telah lampau (Witten et al. 2011). Teknik klasifikasi banyak diterapkan dalam dunia nyata seperti halnya dalam dunia pendidikan (Kotsiantis et al. 2003) (Hastuti 2012) (Ragab et al. 2014), medis (Christobel and Sivaprakasam 2011), teknik bangunan (Ashari et al. 2013).

Klasifikasi *Naive Bayes* merupakan salah satu algoritma dalam teknik data *mining* yang menerapkan teori *Bayes* dalam klasifikasi (Santosa, 2007). Definisi lain mengatakan *Naive Bayes* merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu teorema yang digunakan dalam statistika untuk memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya (Bustami, 2013). Berdasarkan penelitian terkait oleh Bisri (2015) mengenai implementasi Algoritma *Naive Bayes* untuk memprediksi penjurusan siswa di SMA Kesatrian 1 Semarang. Dalam penelitian ini penggunaan nilai mata pelajaran, minat, dan jurusan sekolah menengah sebagai kriteria, diperoleh hasil akurasi sebesar 83.8798 %. Penelitian lain yang dilakukan oleh Hailitik, Djahi dan Nabuasa (2017) mengenai Klasifikasi Jurusan Menggunakan Metode *Naive Bayes* Pada

Sekolah Menengah Atas Negeri (SMAN) 1 Fatuleu Tengah menunjukkan hasil akurasi sebesar 99.31 %. Adapun Penelitian lain yang juga dilakukan oleh Arif Sofanudin (2017) mengenai implementasi Algoritma *Naive Bayes* untuk penjurusan di Ma Al Mahrusiyah Lirboyo dapat diselesaikan dengan menggunakan algoritma *Naive Bayes* untuk membantu pihak Ma dalam proses penentuan jurusan siswa. Keuntungan penggunaan *Naive Bayes* adalah bahwa metode ini hanya membutuhkan jumlah data pelatihan (*training data*) yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian. Akan tetapi performa algoritma ini akan menurun jika terdapat *imbalanced data* dan antar atribut tidak memiliki keterkaitan satu sama lain.

Ada beberapa teknik yang bisa digunakan untuk meningkatkan akurasi dari klasifikasi *Naive Bayes* tersebut, yaitu dengan menggunakan *SMOTE* yang berfungsi untuk menangani ketidakseimbangan kelas dan uji *Chi Square* sebagai fitur seleksi. *Synthetic Minority Oversampling Technique* (SMOTE) adalah salah satu turunan dari *oversampling*. *SMOTE* pertama kali diperkenalkan oleh Nithes V. Chawla. Pendekatan ini bekerja dengan membuat replikasi dari data minoritas. Replikasi tersebut dikenal dengan data sintetis (*syntetic data*). Penerapan *SMOTE* yaitu dengan meminimalisasi keseimbangan kelas sehingga diharapkan dapat memiliki model yang baik (Siringoringo, 2018). *Chi Square Test* (Uji Chi Kuadrat) adalah latihan statistik yang membandingkan frekuensi berbagai jenis atau kategori item dalam sampel acak dengan frekuensi yang diharapkan jika populasi frekuensi seperti yang dihipotesiskan. *Chi Square* digunakan untuk uji independensi dan estimasi yang bertujuan untuk mengetahui ketergantungan suatu *class* pada suatu fitur (Nachirat R, Wattana P, 2015).

Berdasarkan latar belakang tersebut, penelitian ini menerapkan algoritma *Naive Bayes* untuk menghasilkan model yang dapat memprediksi bidang keahlian mahasiswa. *Preprocessing* menggunakan *Synthetic Minority Oversampling Technique* (SMOTE) dan seleksi fitur menggunakan *Chi Square*. Data yang digunakan dalam penelitian ini adalah data survei berupa jawaban kuesioner kemampuan yang mahasiswa kuasai terhadap kriteria sub kemampuan yang dimiliki di setiap kelompok keahlian yang ditujukan kepada mahasiswa

Informatika Fakultas Teknik Universitas Tanjungpura angkatan 2016, 2017 dan 2018. Dari penelitian tersebut, model yang dihasilkan diharapkan dapat memprediksi dan mempermudah dalam mengetahui bidang keahlian mahasiswa berdasarkan kemampuan yang dikuasai.

1.2 Perumusan Masalah

Berdasarkan latar belakang permasalahan, rumusan masalah pada penelitian ini adalah bagaimana mengimplementasikan algoritma *Naive Bayes* untuk menghasilkan model yang dapat memprediksi bidang keahlian mahasiswa Informatika Fakultas Teknik Universitas Tanjungpura.

1.3 Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah menghasilkan model yang dapat memprediksi bidang keahlian mahasiswa dengan menggunakan algoritma *Naive Bayes*.

1.4 Pembatasan Masalah

Peneliti memberikan batasan masalah diberikan untuk menghindari adanya penyimpangan maupun pelebaran inti masalah dalam penyusunan penelitian ini, yaitu:

1. Data yang digunakan adalah data survei berupa jawaban kuesioner kemampuan yang mahasiswa kuasai terhadap kemampuan yang dimiliki kelompok keahlian yang ditujukan kepada mahasiswa Jurusan Informatika Fakultas Teknik Universitas Tanjungpura angkatan 2016, 2017 dan 2018.
2. Memprediksi bidang keahlian mahasiswa berdasarkan kemampuan yang dikuasai.
3. Melakukan pengujian dan mengevaluasi hasil pengujian klasifikasi dengan menggunakan metode *K-Fold Cross Validation* dan *Confusion Matrix*.

1.5 Sistematika Penulisan

Sistematika laporan tugas akhir ini disusun dalam 5 (lima) bab yang terdiri dari Bab I Pendahuluan, Bab II Tinjauan Pustaka, Bab III Metodologi Penelitian, Bab IV Hasil dan Analisis, dan Bab V Penutup.

Bab I Pendahuluan adalah bab yang berisi latar belakang, perumusan masalah, tujuan penelitian, pembatasan masalah, dan sistematika penulisan laporan tugas akhir.

Bab II tinjauan pustaka adalah bab yang membahas mengenai gambaran umum tentang penelitian yang didapat oleh peneliti sebelumnya serta perangkat lunak yang melandasi pembangunan sistem dan landasan teori yang berhubungan dalam proses analisis permasalahan penelitian yang akan dilakukan.

Bab III metodologi penelitian adalah bab yang membahas mengenai data dan perangkat penelitian, metode yang akan digunakan pada penelitian, diagram alir penelitian, dan perancangan sistem.

Bab IV hasil dan analisis adalah bab yang berisi tentang hasil perancangan, pengoperasian sistem, pengujian dan analisis pengujian yang mengarah kepada suatu kesimpulan.

Bab V penutup adalah bab yang berisi kesimpulan dari penelitian yang telah dilakukan dan saran atau rekomendasi untuk perbaikan maupun pengembangan serta kelengkapan penelitian.

BAB II

TINJAUAN PUSTAKA

2.1 Kajian Terkait

Pada penelitian terkait ini, adalah proses mengkaji penelitian mengenai klasifikasi yang dilakukan oleh peneliti-peneliti sebelumnya yang dapat menjadi dasar untuk penelitian yang dilakukan. Beberapa kajian terkait dengan penelitian ini diantaranya sebagai berikut:

Swastina (2013) melakukan penelitian dengan judul Penerapan Algoritma C4.5 Untuk Penentuan Jurusan Mahasiswa. Penelitian ini melakukan klasifikasi terhadap mahasiswa dengan pemilihan konsentrasi ditentukan berdasarkan kriteria minat, dan indeks Prestasi Kumulatif Semester 1 dan 2. Penelitian ini bertujuan untuk menentukan jurusan yang akan diambil sesuai dengan minat dan kemampuannya.

Bisri (2015) melakukan penelitian dengan judul Implementasi Algoritma *Naive Bayes* untuk Memprediksi Penjurusan Siswa di SMA Kesatrian 1 Semarang. Penelitian ini melakukan penjurusan bagi siswa berdasarkan prestasi siswa yang akan memasuki jenjang pendidikan menengah atas. Komponen kriteria yang dibutuhkan untuk mendapatkan rekomendasi keputusan dalam penelitian ini adalah minat, nilai mata pelajaran. Penelitian ini bertujuan untuk memprediksi jurusan siswa SMA Kesatrian 1 Semarang.

Sofanudin (2017) melakukan penelitian dengan judul Implementasi Algoritma *Naive Bayes* Untuk Penjurusan Di Ma Al Mahrusiyah Lirboyo. Penelitian ini melakukan pengelompokan data siswa dengan menggunakan metode implementasi *Naive Bayes* dan program Php yang bertujuan untuk menentukan jurusan pada siswa. Kriteria yang dibutuhkan dalam penelitian ini adalah minat, nilai mata pelajaran, dan jurusan yang ada di Ma Al Mahrusiyah Lirboyo.

Hailitik, Djahi, Nabuasa (2017) melakukan penelitian dengan judul Klasifikasi Jurusan Menggunakan Metode *Naive Bayes* Pada Sekolah Menengah

Atas Negeri (SMAN) 1 Fatuleu Tengah. Penelitian ini melakukan klasifikasi jurusan bagi siswa, penjurusan pada SMA Negeri 1 Fatuleu Tengah terdiri dari IPA dan IPS. Data siswa yang digunakan dalam penelitian ini adalah data siswa SMAN 1 Fatuleu Tengah kelas XI semester 2 yaitu data nilai rapor (matematika, fisika, kimia, biologi, geografi, ekonomi, sosiologi, sejarah), data nilai bakat (nilai IQ, *verbal*, *numeral*, *spasial*, *persepsional*, teknik) dan data nilai minat (minat orang tua dan minat siswa).

Faidhani (2018) melakukan penelitian dengan judul Sistem Pendukung Keputusan Penentu Bidang Keahlian Mahasiswa Program Studi Teknik Informatika Universitas Tanjungpura Dengan Metode ELECTRE. Pengujian yang ditentukan berdasarkan rata-rata nilai mata kuliah kelompok keahlian, minat mahasiswa, dan kemampuan mahasiswa (*hard skill*). Tujuan dari penelitian yang dilakukan adalah menghasilkan rekomendasi dalam menentukan bidang keahlian mahasiswa di Program Studi Informatika Universitas Tanjungpura.

Istighfar (2020) dengan penelitian yang berjudul Klasifikasi Bidang Keahlian Mahasiswa Menggunakan Algoritma *Naive Bayes*. Penelitian ini melakukan pengelompokan data mahasiswa dengan menggunakan metode implementasi *Naive Bayes*, *SMOTE* dan *feature selection Chi Square* yang bertujuan menghasilkan model yang dapat memprediksi bidang keahlian mahasiswa. Kriteria yang dibutuhkan dalam penelitian ini adalah penguasaan sub kemampuan yang dikuasai oleh mahasiswa terhadap kriteria sub kemampuan yang ada di Jurusan Informatika Fakultas Teknik Universitas Tanjungpura.

Perbedaan dari penelitian yang dilakukan pada penelitian sebelumnya dapat dilihat pada Tabel 2.

Tabel 2. 1 Tabel Perbandingan Penelitian

No.	Penulis	Judul Penelitian	Keterangan
1.	Liliana Swastina (2013)	Penerapan Algoritma C4.5 Untuk Penentuan Jurusan Mahasiswa.	Kriteria yang digunakan adalah minat, dan indeks Prestasi Kumulatif Semester 1 dan 2.
2.	Muchamad Bisri Hasan (2015)	Implementasi Algoritma <i>Naive Bayes</i> untuk Memprediksi Penjurusan Siswa di SMA Kesatrian 1 Semarang.	Menggunakan nilai mata pelajaran, minat, dan jurusan sekolah menengah sebagai kriteria.
3.	Arif Sofanudin (2017)	Implementasi Algoritma <i>Naive Bayes</i> Untuk Penjurusan Di Ma Al Mahrusiyah Lirboyo.	Kriteria yang dibutuhkan dalam penelitian ini adalah minat, nilai mata pelajaran, dan jurusan di Ma Al Mahrusiyah Lirboyo.
4.	Arddy H. Hailitik, Bertha S. Djahi, Yelly Y. Nabuasa (2017)	Klasifikasi Jurusan Menggunakan Metode <i>Naive Bayes</i> Pada Sekolah Menengah Atas Negeri (Sman) 1 Fatuleu Tengah.	Data siswa yang digunakan dalam penelitian ini adalah data siswa SMAN 1 Fatuleu Tengah kelas XI semester 2 yaitu data nilai rapor (matematika, fisika, kimia, biologi, geografi, ekonomi, sosiologi, sejarah), data nilai bakat (nilai IQ, verbal, numeral, spasial, persepsional, teknik) dan data nilai minat (minat orangtua dan minat siswa).

No.	Penulis	Judul Penelitian	Keterangan
5.	Faz Faidhani (2018)	Sistem Pendukung Keputusan Penentu Bidang Keahlian Mahasiswa Program Studi Teknik Informatika Universitas Tanjungpura Dengan Metode ELECTRE.	Kriteria yang digunakan adalah rata-rata nilai mata kuliah kelompok keahlian, minat mahasiswa, dan kemampuan mahasiswa (hard skill).
6.	Fajar Istighfar (2020)	Klasifikasi Bidang Keahlian Mahasiswa Menggunakan Algoritma <i>Naive Bayes</i> .	Kriteria yang digunakan dalam penelitian ini adalah penguasaan sub kemampuan yang dikuasai oleh mahasiswa terhadap kriteria sub kemampuan yang ada di Jurusan Informatika Fakultas Teknik Universitas Tanjungpura.

2.2 Jurusan Informatika Fakultas Teknik Universitas Tanjungpura

Program Studi Informatika merupakan salah satu program studi jenjang sarjana (S1) di lingkungan Fakultas Teknik Universitas Tanjungpura, berdiri pada tanggal 18 Mei 2014 dengan SK Dirjen DIKTI Nomor 1664/D/T/2004. Saat ini penyelenggaraan Program Studi Informatika berdasarkan SK Penyelenggaraan Nomor 7325/D/T/K-N/2011 tanggal 6 Juni 2011, dan Terakreditasi B sesuai SKBAN-PT Nomor 0770/SK/BAN-PT/Akred/S/III/2017 pada 21 Maret 2017 (Teknik Informatika, 2016).

Penyelenggaraan pendidikan, Program Studi Informatika melaksanakan dua program studi pendidikan, yaitu program S1 Reguler A dan program S1

Reguler B. Kedua program ini melaksanakan kurikulum pendidikan yang sama, hanya dibedakan waktu pelaksanaan. Program Studi Informatika hingga saat ini masih menggunakan Kurikulum Kerangka Kualifikasi Nasional Indonesia (KKNI) tahun 2016. Program Studi Informatika Universitas Tanjungpura mempersiapkan lulusannya untuk terjun di masyarakat sesuai dengan standar kompetensi kerja baik nasional maupun internasional, khususnya bidang Informatika. Program Studi Informatika mensyaratkan pembelajaran baik teori untuk penguasaan pengetahuan, praktik di laboratorium/lapangan untuk penguasaan kemampuan kerja (*skills*), maupun kecakapan hidup dan kemampuan manajerial (Teknik Informatika, 2016).

Program Studi Informatika Universitas Tanjungpura memiliki 4 kelompok keahlian sebagai berikut:

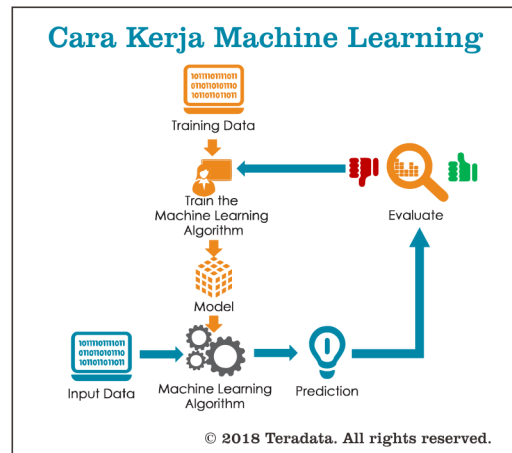
1. Sistem Informasi dan Data Spasial
2. Rekayasa Perangkat Lunak dan Komputasi Bergerak
3. Jaringan dan Keamanan
4. Komputasi dan Kecerdasan Buatan

2.3 *Machine Learning*

Machine Learning adalah aplikasi dari disiplin ilmu kecerdasan buatan (*Artificial Intelligence*) yang menggunakan teknik statistika untuk menghasilkan suatu model otomatis dari sekumpulan data, dengan tujuan memberikan komputer kemampuan untuk "belajar". Pembelajaran mesin atau *machine learning* memungkinkan komputer mempelajari sejumlah data (*learn from data*) sehingga dapat menghasilkan suatu model untuk melakukan proses *input-output* tanpa menggunakan kode program yang dibuat secara eksplisit. Proses belajar tersebut menggunakan algoritma khusus yang disebut *machine learning algorithms*. Terdapat banyak algoritma *machine learning* dengan efisiensi dan spesifikasi kasus yang berbeda-beda (Ralf Herbrich and Graepel, 2018).

Secara fundamental cara kerja *machine learning* adalah belajar seperti manusia dengan menggunakan contoh-contoh dan setelah itu barulah dapat

menjawab suatu pertanyaan terkait. Proses belajar ini menggunakan data yang disebut *train data set*. Berbeda dengan program statis, machine learning diciptakan untuk membentuk program yang dapat belajar sendiri. Cara kerja *machine learning* dapat dilihat pada gambar 2.1



Gambar 2. 1 Cara kerja *machine learning*

Dari data tersebut, komputer akan melakukan proses belajar (*training*) untuk menghasilkan suatu model. Proses belajar ini menggunakan algoritma *machine learning* sebagai penerapan teknik statistika. Model inilah yang menghasilkan informasi, kemudian dapat dijadikan pengetahuan untuk memecahkan suatu permasalahan sebagai proses *input-output*. Model yang dihasilkan dapat melakukan klasifikasi atau pun prediksi kedepannya.

Untuk memastikan efisiensi model yang terbentuk, data akan dibagi menjadi data pembelajaran (***train data set***) dan data pengujian (*test data set*). Pembagian data yang digunakan bervariasi bergantung algoritma yang digunakan. Pada umumnya *train data set* lebih banyak dari ***test data set***, misalnya dengan rasio 3:1. *Test data set* digunakan untuk menghitung seberapa efisien model yang dihasilkan untuk melakukan klasifikasi atau prediksi kedepannya yang disebut ***test score***. Semakin banyak data yang digunakan, *test score* yang dihasilkan semakin baik. Nilai *test score* berada dalam rentang 0-1.

Metode Algoritma *Machine Learning*.

1. Supervised machine learning algorithms

Supervised machine learning adalah algoritma *machine learning* yang dapat menerapkan informasi yang telah ada pada data dengan memberikan label tertentu, misalnya data yang telah diklasifikasikan sebelumnya (terarah). Algoritma ini mampu memberikan target terhadap *output* yang dilakukan dengan membandingkan pengalaman belajar di masa lalu.

2. Unsupervised machine learning algorithms

Unsupervised machine learning adalah algoritma *machine learning* yang digunakan pada data yang tidak mempunyai informasi yang dapat diterapkan secara langsung (tidak terarah). Algoritma ini diharapkan mampu menemukan struktur tersembunyi pada data yang tidak berlabel.

3. Semi-supervised machine learning algorithms

Semi-supervised machine learning adalah algoritma yang digunakan untuk melakukan pembelajaran data berlabel dan tanpa label. Sistem yang menggunakan metode ini dapat meningkatkan efisiensi *output* yang dihasilkan.

4. Reinforcement machine learning algorithms

Reinforcement machine learning adalah algoritma yang mempunyai kemampuan untuk berinteraksi dengan proses belajar yang dilakukan, algoritma ini akan memberikan poin (*reward*) saat model yang diberikan semakin baik atau mengurangi poin (*error*) saat model yang dihasilkan semakin buruk. Salah satu penerapannya adalah pada mesin pencari.

2.4 Klasifikasi

Klasifikasi adalah proses pengidentifikasian objek ke dalam sebuah kategori, kelas atau kelompok berdasarkan prosedur, definisi dan karakteristik yang telah ditentukan sebelumnya (U.S Fish and Wildlife Service, 2013).

Klasifikasi bertujuan untuk menempatkan objek yang ditugaskan hanya ke salah satu kategori yang disebut kelas (Bramer, 2007). Tak hanya klasifikasi, proses pengelompokan objek juga dapat dilakukan dengan menggunakan teknik *clustering*. *Clustering* merupakan pengelompokan obyek berdasarkan kemiripan antar objek. Perbedaan antara klasifikasi dan *clustering* terletak pada proses pengelompokan objek. Jika pada klasifikasi proses pengelompokan obyek dilakukan dengan membagi obyek berdasarkan kelompok / kategori yang telah didefinisikan sebelumnya, maka proses pengelompokan objek pada *clustering* dilakukan dengan melihat kemiripan antar obyek, sehingga kategori belum terdefinisi sebelumnya. Salah satu metode klasifikasi yang sering digunakan adalah *Naive Bayes Classifier* yang pertama kali dikemukakan oleh Revered Thomas Bayes. Penggunaan *Naive Bayes Classifier* sudah dikenalkan sejak tahun 1702-1761. Menurut Lewis (1998), Hand dan Yu (2001), *Naive Bayes Classifier* merupakan pendekatan yang sangat sederhana dan sangat efektif untuk pelatihan klasifikasi. Sedangkan Kononenko dan Langley menyimpulkan bahwa *Naive Bayes Classifier* merupakan kemungkinan label kelas data atau bisa diasumsikan sebagai atribut kelas yang diberi label (Kononenko, 1990) (Langley, 1994).

2.5 Bahasa Pemrograman Python

Python adalah bahasa pemrograman interpretatif multiguna. Tidak seperti bahasa lain yang susah untuk dibaca dan dipahami, python lebih menekankan pada keterbacaan kode agar lebih mudah untuk memahami sintaks. Hal ini membuat Python sangat mudah dipelajari baik untuk pemula maupun untuk yang sudah menguasai bahasa pemrograman lain. Bahasa ini muncul pertama kali pada tahun 1991, dirancang oleh seorang bernama Guido van Rossum. Sampai saat ini Python masih dikembangkan oleh Python Software Foundation. Bahasa Python mendukung hampir semua sistem operasi, bahkan untuk sistem operasi Linux, hampir semua distronya sudah menyertakan Python di dalamnya. Dengan kode yang simpel dan mudah diimplementasikan, seorang programmer dapat lebih mengutamakan pengembangan aplikasi yang dibuat, bukan malah sibuk mencari *syntax error*.

2.6 Jupyter Notebook

Jupyter adalah singkatan dari tiga bahasa pemrograman **Julia** (Ju), **Python** (Py), dan **R**. Tiga bahasa pemrograman ini adalah sesuatu yang penting bagi seorang *data scientist*. Jupyter berfungsi untuk membantu dalam membuat narasi komputasi yang menjelaskan makna dari data di dalamnya dan memberikan *insight* mengenai data tersebut. Selain itu, Jupyter juga mempermudah kerja sama antara insinyur dan *data scientist* karena kemudahannya dalam menulis dan berbagi teks dan kode. Karena alasan inilah, Jupyter mempermudah *data scientist* untuk berkolaborasi dengan *data scientist*, *data researchers* atau *data engineers* lainnya.

2.7 Scikit-Learn

Scikit-Learn merupakan *library Machine Learning open source* berbasis Python yang bisa digunakan dalam *Data Science*. Kelebihan *Scikit-Learn* adalah penggunaan API yang mudah serta kecepatannya saat melakukan tolak ukur yang berbeda dalam data set. *Sklearn* kompatibel dengan *NumPy* dan *SciPy*. Ini artinya kamu akan dapat beroperasi dengan *library-library* yang berbeda untuk Python dengan mudah. *Scikit-Learn* memberikan sejumlah fitur untuk keperluan *Data Science* seperti algoritma *Regresi*, pengelompokan, algoritma *Naive Bayes*, algoritma *Decision Tree*, *parameter tuning*, *data preprocessing tool*, *export/import model*, *Machine learning pipeline* dan algoritma klasifikasi termasuk gradien, *K-means*, mesin dukungan vektor, DBSCAN, dan juga mampu beroperasi dengan *SciPy* dan *NumPy*.

2.8 Algoritma Naive Bayes

Menurut Mochammad Haldi Widiyanto (2019) Algoritma *Naive Bayes* merupakan salah satu algoritma yang terdapat pada teknik klasifikasi. *Naive Bayes* merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman dimasa sebelumnya dengan menghitung peluang dari satu kelas dari masing-masing kelompok atribut yang ada, dan menentukan kelas mana yang paling optimal, sehingga dikenal sebagai Teorema

Bayes. Teorema tersebut dikombinasikan dengan *Naive* dimana diasumsikan kondisi antar atribut saling bebas. Proses pengelompokan atau klasifikasi dibagi menjadi dua proses yaitu *learning/training* dan *testing/classify*. Pada proses *learning*, sebagian data yang telah diketahui kategorinya, datanya diumpankan untuk membentuk model perkiraan. Kemudian pada proses *testing*, model yang sudah terbentuk diuji dengan sebagian data.

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \quad (2.1)$$

Keterangan :

X = data dengan *class* yang belum diketahui

H = hipotesis data X, merupakan suatu class spesifik $P(H|X)$ = probabilitas hipotesis H berdasarkan kondisi X (*posteriori probability*)

$P(H)$ = probabilitas hipotesis H (*prior probability*)

$P(X|H)$ = probabilitas X berdasarkan kondisi pada hipotesis H

$P(X)$ =probabilitas dari X

2.9 Bernoulli Naive Bayes

Bernoulli Naive Bayes adalah algoritma *Naive Bayes* yang mengasumsikan beberapa fitur ke dalam *binary*, jadi model *Bernoulli Naive Bayes* ini menangkap frekuensi kata dengan menggunakan *boolean*, jika kata tersebut terdapat dalam dokumen maka bernilai 1 dan jika kata tersebut tidak ada dalam dokumen maka akan bernilai 0.

$$P(t | c) = \frac{P(Tct)+Smoothing}{(\sum Tct)t \in V + |V|} \quad (2.2)$$

Keterangan:

$P(Tct)$ = jumlah istilah yang ada di kelas

Smoothing = di tambahkan dengan 1 untuk menghindari hasil 0

$(\sum Tct)t \in V$ = jumlah kelas yang ada pada dokumen

$|V|$ = jumlah prediksi yang ada pada suatu istilah

2.10 *Synthetic Minority Oversampling Technique (SMOTE)*

Akurasi dari kelas data sangat penting dalam klasifikasi, semakin akurat data set dan kelas yang ada, semakin baik output yang dihasilkan. Masalah imbalanced data dapat menyebabkan hasil dari klasifikasi tidak akurat sehingga dibutuhkan teknik agar data set yang dimiliki menjadi seimbang sehingga dapat menghasilkan data yang akurat. Salah satu cara yang paling mudah untuk memperbaiki data yang tidak seimbang adalah dengan membuat data tersebut menjadi seimbang adalah *Oversampling* pada *minority* kelas atau *Undersampling* pada *majority class*. Dengan metode *Over/under sampling* dengan mudah dapat membuat data-set menjadi seimbang tetapi metode ini mempunyai kelemahan, *Oversampling* pada data-set *minority* akan menuju model yang *overfitting*, karena *Oversampling* dilakukan dengan duplikasi data yang sudah mempunyai nilai yang sudah kecil, *Undersampling* pada *majority* juga dapat mengakibatkan data yang penting pembeda dua kelas menjadi diluar dari data-set. Metode *Synthetic Minority Over-sampling Technique (SMOTE)* merupakan metode yang diterapkan dalam penelitian ini untuk menangani ketidakseimbangan kelas (Arwan et al., 2018)

2.11 *Chi Square*

Chi Square atau Kai Kuadrat adalah salah satu jenis uji komparatif non parametris yang dilakukan pada dua variabel, dimana skala data kedua variabel adalah nominal (Sutrisno,2000). Apabila dari 2 variabel, ada 1 variabel dengan skala nominal maka dilakukan uji *Chi Square* dengan merujuk bahwa harus digunakan uji pada derajat yang terendah. *Chi Square* memiliki 3 uji, yaitu *observe frequency*, *expected frequency*, dan *test statistic*. Syarat uji ini adalah frekuensi responden atau sampel yang digunakan adalah besar, berikut syarat uji *Chi Square* yaitu :

1. tidak ada sel dengan nilai frekuensi kenyataan atau disebut juga *Actual Count* (F_0) sebesar 0 (Nol).
2. apabila bentuk tabel kontingensi 2 X 2, maka tidak boleh ada 1 sel saja yang memiliki frekuensi harapan atau disebut juga *expected count* (“ F_h ”) kurang dari 5.
3. apabila bentuk tabel lebih dari 2 x 2, misal 2 x 3, maka jumlah sel dengan frekuensi harapan yang kurang dari 5 tidak boleh lebih dari 20%.

2.12 K-Fold Cross Validation

Menurut Yaya Wihardi (2013) *Cross Validation* merupakan salah satu teknik untuk menilai/memvalidasi keakuratan sebuah model yang dibangun berdasarkan data set tertentu. Pembuatan model biasanya bertujuan untuk melakukan prediksi maupun klasifikasi terhadap suatu data baru yang boleh jadi belum pernah muncul di dalam data set. Data yang digunakan dalam proses pembangunan model disebut data latih/*training*, sedangkan data yang akan digunakan untuk memvalidasi model disebut sebagai *data test*.

Salah satu metode *cross-validation* yang populer adalah *K-Fold Cross Validation*. Dalam teknik ini data set dibagi menjadi sejumlah K-buah partisi secara acak. Kemudian dilakukan sejumlah K-kali eksperimen, dimana masing-masing eksperimen menggunakan data partisi ke-K sebagai data *testing* dan memanfaatkan sisa partisi lainnya sebagai data *training*.

2.13 Pengujian Confusion Matrix

Confusion Matrix adalah evaluasi dari sebuah klasifikasi data *mining* yang direpresentasikan menjadi tabel (Gorunescu, 2011). *Confusion Matrix* berisi informasi perbandingan label hasil klasifikasi dengan label sebenarnya (Indriyanti, Sugianti and Karomi, 2017). Pada pengukuran kinerja menggunakan *Confusion Matrix*, terdapat 4 (empat) istilah sebagai representasi hasil proses klasifikasi. Keempat istilah tersebut adalah:

1. *True Positive* (TP) : data positif yang terdeteksi benar

2. *True Negative* (TN) : data negatif yang terdeteksi dengan benar
3. *False Positive* (FP) : data negatif namun terdeteksi sebagai data positif
4. *False Negative* (FN) : data positif namun terdeteksi sebagai data negatif

Dalam dunia pengenalan pola (*pattern recognition*) dan pencarian informasi (*information retrieval*), *precision* dan *recall* adalah perhitungan yang banyak digunakan untuk mengukur kinerja dari sistem atau metode yang digunakan. *Precision* adalah tingkat ketepatan antara informasi yang diminta oleh pengguna dengan jawaban yang diberikan oleh sistem. *Recall* adalah persentase keberhasilan sistem dalam menemukan kembali sebuah informasi. *Accuracy* adalah tingkat kedekatan antara nilai prediksi dan nilai aktual (Han et al., 2012). Sedangkan *F1-Score* didefinisikan sebagai rata-rata harmonis atau rata-rata tertimbang dari *precision* dan *recall*.

Tabel 2. 2 *Confusion Matrix*

		Nilai Sebenarnya	
		TRUE	FALSE
Nilai Prediksi	TRUE	TP (True Positive) Correct result	FP (False Positive) Unexpected result
	FALSE	FN (False Negative) Missing result	TN (True Negative) Correct absence of result

$$Precision = \frac{TP}{TP+FP} \quad (2.3)$$

$$Recall = \frac{TP}{TP+FN} \quad (2.4)$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (2.5)$$

$$F1-Score = \frac{2 \times Recall \times Precision}{Recall + Precision} \quad (2.6)$$

BAB III METODOLOGI PENELITIAN

3.1 Bahan Penelitian

Bahan penelitian adalah data survei berupa jawaban kuesioner kemampuan yang dikuasai oleh mahasiswa terhadap kemampuan yang dimiliki di setiap kelompok keahlian Jurusan Informatika Fakultas Teknik Universitas Tanjungpura yang ditujukan kepada mahasiswa Jurusan Informatika Fakultas Teknik Universitas Tanjungpura angkatan 2016 sampai 2018 bagi yang sedang mengajukan skripsi maupun yang sudah selesai skripsi.

3.2 Perangkat Pendukung Penelitian

3.2.1 Perangkat Lunak

Perangkat lunak yang digunakan dalam penelitian ini adalah:

1. Sistem Operasi *Windows* 10

Sistem Operasi *Windows* berfungsi untuk mengendalikan kerja perangkat keras seperti *CPU*, *harddisk*, monitor dan lain-lain serta mengatur fungsi program *software* agar terhubung dengan perangkat keras tersebut.

2. *Scikit Learn*

Scikit-learn adalah *library* untuk *machine learning* bagi para pengguna python. *Scikit-learn* merupakan *free software*, dan memungkinkan kita melakukan beragam pekerjaan dalam *Data Science*, seperti regresi (*regression*), klasifikasi (*classification*), pengelompokan atau penggugusan (*clustering*), data *preprocessing*, *dimensionality reduction*, dan *model selection* (pembandingan, validasi, dan pemilihan parameter maupun model).

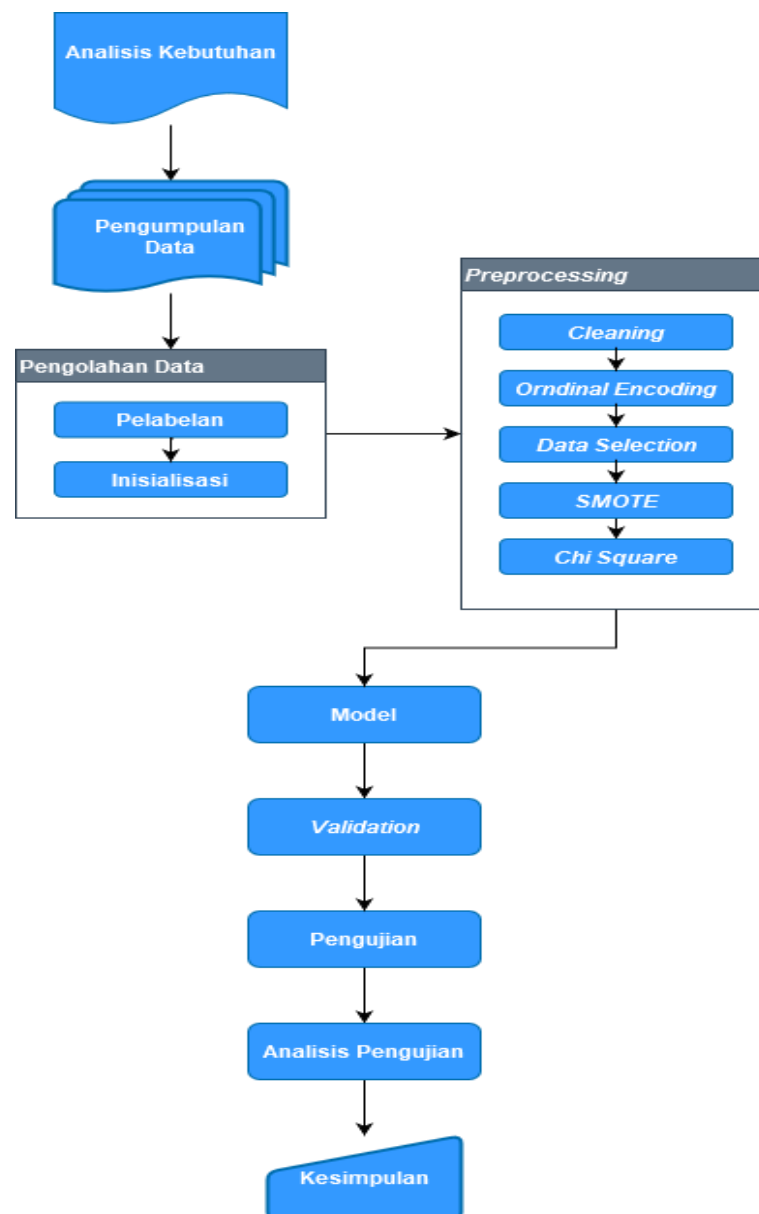
3.2.2 Perangkat Keras

Perangkat Keras yang digunakan dalam penelitian ini adalah:

PC/Laptop ASUS X441U dengan spesifikasi prosesor Intel Core i3-6006U, 2.0GHz, RAM 4 GB, HDD 1 TB.

3.3 Metode Penelitian

Langkah-langkah penelitian yang dilakukan seperti terlihat pada Gambar 3.1 sebagai berikut :



Gambar 3. 1 Metode penelitian

Berikut penjelasan beberapa tahapan dalam penelitian yang memenuhi tugas akhir ini adalah sebagai berikut:

1. Analisis Kebutuhan

Analisis kebutuhan merupakan tahapan menganalisa dan memahami kebutuhan pada penelitian yang akan dilakukan melalui wawancara dan observasi terhadap ketua dan anggota kelompok keahlian di Jurusan Informatika Fakultas Teknik Universitas Tanjungpura untuk mendapatkan *list* kemampuan yang diperlukan di setiap kelompok keahlian.

2. Pengumpulan Data

Data merupakan data survei berupa jawaban kuesioner kemampuan yang dikuasai mahasiswa terhadap kemampuan yang dimiliki di setiap kelompok keahlian yang ditujukan kepada mahasiswa Jurusan Informatika Fakultas Teknik Universitas Tanjungpura angkatan 2016 sampai 2018 bagi yang sedang mengajukan skripsi maupun yang sudah selesai skripsi.

3. Pengolahan Data

Pelabelan data adalah proses mengidentifikasi data mentah (gambar, *file teks*, video, dll.) dan menambahkan satu atau lebih label yang bermakna dan informatif untuk memberikan konteks sehingga model pembelajaran mesin dapat belajar darinya. Pelabelan data diperlukan untuk berbagai kasus penggunaan termasuk visi komputer, pemrosesan bahasa alami, dan pengenalan suara. Inisialisasi adalah pemberian nilai awal yang dilakukan saat deklarasi variable atau objek.

4. *Preprocessing*

Tahap ini menggunakan proses *Cleaning*, *Ordinal Encoding*, *Data Selection* dan *Synthetic Minority Over-sampling Technique* (SMOTE). Data terlebih dahulu harus diperiksa dari data yang inkonsisten, diperbaiki kesalahan tipografi, data yang berupa kategorikal diubah menjadi data *numerical*, selanjutnya data yang tidak akan digunakan dan memiliki ketidakseimbangan, di seleksi setelah itu dilakukan teknik penyeimbangan pada data dan kelas yang ada hingga menjadi data akhir atau data set yang dapat diproses di *Machine Learning*.

5. *Feature Selection*

Tahapan ini merupakan proses meminimalkan data yang akan digunakan untuk proses mencari dan menentukan kemampuan apa saja yang paling relevan terhadap kelompok keahlian untuk dijadikan fitur dengan menggunakan *Chi Square*.

6. *Pemodelan*

Tahap utama dari penelitian ini adalah klasifikasi dengan menggunakan algoritma *Naive Bayes*. Pada tahap ini akan memanggil fungsi *Naive Bayes* dari library Scikit Learn untuk melakukan proses *training* pada data yang telah melalui tahapan preprocessing dan *feature selection*.

7. *Validation*

Pada Tahap ini data set dibagi menjadi sejumlah K-buah partisi secara acak, kemudian dilakukan sejumlah K-kali eksperimen, dimana masing-masing eksperimen menggunakan data partisi ke-K sebagai data testing dan memanfaatkan sisa partisi lainnya sebagai data training.

8. *Pengujian*

Penelitian ini menggunakan *Scikit Learn* untuk mengolah data dan melakukan pengujian dengan menggunakan pengujian *Confusion Matrix* sebagai alat bantu dalam mengukur akurasi data eksperimen yang dilakukan dalam penelitian

9. *Analisis Hasil Pengujian*

Analisis hasil pengujian dilakukan untuk mengetahui apakah model yang dibuat telah menghasilkan hasil yang sesuai dengan perancangan yang sudah ditentukan. Pada hasil analisis pengujian dapat digunakan sebagai dasar penarikan kesimpulan.

10. *Kesimpulan*

Kesimpulan dirumuskan berdasarkan hasil pengujian yang telah dilakukan apakah model berhasil mengklasifikasikan mahasiswa ke dalam kelompok keahlian menggunakan metode Klasifikasi *Naive Bayes*.

3.3.1 Analisis Kebutuhan

Analisis kebutuhan adalah tahapan menganalisa dan memahami kebutuhan pada penelitian yang akan dilakukan melalui wawancara dan observasi terhadap ketua dan anggota kelompok keahlian di Jurusan Informatika Fakultas Teknik Universitas Tanjungpura untuk mendapatkan list kriteria sub kemampuan yang dimiliki di setiap kelompok keahlian, list kriteria sub kemampuan dapat dilihat pada tabel 3.1. Adapun kemampuan yang dimiliki terbagi menjadi:

1. Penguasaan Bahasa Pemrograman
2. Penguasaan *Tools*/Aplikasi
3. Pemahaman dan Penerapan Konsep

Tabel 3. 1 Kriteria Sub Kemampuan

Kelompok Keahlian Sistem Informasi dan Data Spasial	Kelompok Keahlian Rekayasa Perangkat Lunak dan Komputasi Bergerak	Kelompok Keahlian Jaringan dan Keamanan	Kelompok Keahlian Komputasi dan Kecerdasan Buatan
1. Penguasaan Bahasa Pemrograman a. Bahasa pemrograman XML b. Bahasa pemrograman Java c. Bahasa pemrograman Javascript d. Bahasa pemrograman PHP dan ASP e. Bahasa pemrograman HTML f. Bahasa pemrograman SQL	1. Penguasaan Bahasa Pemrograman a. Bahasa pemrograman XML b. Bahasa pemrograman Java c. Bahasa pemrograman Javascript d. Bahasa pemrograman PHP dan ASP e. Bahasa pemrograman HTML f. Bahasa pemrograman SQL g. Bahasa pemrograman Python h. Bahasa pemrograman Visual Basic i. Bahasa pemrograman C atau C++	1. Penguasaan Bahasa Pemrograman a. Bahasa pemrograman Java b. Bahasa pemrograman Javascript c. Bahasa pemrograman PHP dan ASP d. Bahasa pemrograman HTML e. Bahasa pemrograman SQL f. Bahasa pemrograman Python g. Bahasa pemrograman C atau C++	1. Penguasaan Bahasa Pemrograman a. Bahasa pemrograman Java b. Bahasa pemrograman Javascript c. Bahasa pemrograman PHP dan ASP d. Bahasa pemrograman SQL e. Bahasa pemrograman Prolog f. Bahasa pemrograman Python g. Bahasa pemrograman C atau C++
2. Penguasaan Tools/Aplikasi a. Tools Apache Spark b. Tensor Flow c. Ms SQL d. Ms Project e. StarUML f. Packet Tracer g. RAD h. Agile	2. Penguasaan Tools/Aplikasi a. Tools Apache Spark b. Tools BigML c. WEKA d. Ms SQL e. Ms Project f. StarUML g. Netbeans h. Eclipse i. RAD j. Agile k. SQLite	2. Penguasaan Tools/Aplikasi a. Tools Apache Spark b. NET cut, Cain dan Abel, Wireshark c. Ms SQL d. Ms Project e. StarUML f. Packet Tracer g. Winbox h. Netbeans i. Eclipse j. RAD k. SQLite	2. Penguasaan Tools/Aplikasi a. Tools Apache Spark b. Tools Matplotlib c. Tools NLTK d. Tools BigML e. Tools Jupiter f. Tools Moses g. Tools Srilm h. Tools Numpy i. Tools Scipy j. Tools Scikit-Learn k. Tools TensorFlow l. Tools Pentaho m. Tools Tableau n. WEKA o. Ms SQL p. StarUML q. Netbeans r. Eclipse

Kelompok Keahlian Sistem Informasi dan Data Spasial	Kelompok Keahlian Rekayasa Perangkat Lunak dan Komputasi Bergerak	Kelompok Keahlian Jaringan dan Keamanan	Kelompok Keahlian Komputasi dan Kecerdasan Buatan
			s. RAD t. Matlab u. SQLite
3. Pemahaman dan Penerapan Konsep a. Pemahaman konsep XHTML b. Pemahaman konsep Desain Grafis c. Penerapan konsep FTP server dan Web server d. Pemahaman Matlab Basic IP dan Subnetting e. Pemahaman sistem informasi dan jenis sistem informasi : MIS, DSS, dan EIS f. Pemahaman konsep Database g. Pemahaman perancangan ERD dan UML h. Pemahaman perancangan arsitektur DBMS i. Pemahaman arsitektur RISC j. Pemahaman konsep manajemen perangkat lunak I/O k. Pemahaman teknik pengujian Model/Aplikasi l. Pemahaman salah satu atau lebih algoritma pembelajaran mesin atau kecerdasan buatan m. Pemahaman konsep mobile computing n. Memahami dan menguasai Search Engine Optimization (SEO)	3. Pemahaman dan Penerapan Konsep a. Pemahaman konsep XHTML b. Pemahaman konsep Desain Grafis c. Pemahaman sistem informasi dan jenis sistem informasi : MIS, DSS, dan EIS d. Pemahaman konsep Database e. Pemahaman perancangan ERD dan UML f. Pemahaman perancangan arsitektur DBMS g. Pemahaman arsitektur RISC h. Pemahaman konsep manajemen perangkat lunak I/O i. Pemahaman teknik pengujian Model/Aplikasi j. Pemahaman konsep mobile computing k. Memahami dan menguasai Search Engine Optimization (SEO)	3. Pemahaman dan Penerapan Konsep a. Pemahaman konsep Monitor dan administrasi Network Security b. Penerapan konsep FTP server dan Web server c. Pemahaman Basic IP dan Subnetting d. Pemahaman konsep Database e. Pemahaman arsitektur RISC f. Pemahaman teknik pengujian Model/Aplikasi g. Pemahaman konsep mobile computing h. Memahami dan menguasai Search Engine Optimization (SEO)	3. Pemahaman dan Penerapan Konsep a. Pemahaman sistem informasi dan jenis sistem informasi : MIS, DSS, dan EIS b. Pemahaman konsep Database c. Pemahaman perancangan ERD dan UML d. Pemahaman perancangan arsitektur DBMS e. Pemahaman konsep manajemen perangkat lunak I/O f. Pemahaman teknik pengujian Model/Aplikasi g. Pemahaman salah satu atau lebih algoritma pembelajaran mesin atau kecerdasan buatan h. Pemahaman konsep mobile computing

3.3.2 Pengumpulan Data

Data set yang digunakan dalam penelitian ini adalah data survei berupa jawaban kuesioner kemampuan yang mahasiswa kuasai terhadap kriteria sub kemampuan yang dimiliki di setiap kelompok keahlian yang ditujukan kepada mahasiswa Informatika Fakultas Teknik Universitas Tanjungpura angkatan 2016, 2017 dan 2018. Pada penelitian ini, yang menjadi data adalah tingkat kemampuan yang dikuasai oleh mahasiswa terhadap kriteria sub kemampuan yang dimiliki di setiap kelompok keahlian Jurusan Informatika Fakultas Teknik Universitas Tanjungpura. Data set berjumlah 132 data dan disimpan dalam file dengan format .csv. Adapun data tersebut dapat dilihat pada tabel 3.2.

Tabel 3. 2 Data Hasil Survei Kemampuan

Time stamp	Email Address	Nama	Nim	1. Penguasaan Bahasa Pemrograman [a. Bahasa pemrograman XML]	1. Penguasaan Bahasa Pemrograman [b. Bahasa pemrograman Java]	1. Penguasaan Bahasa Pemrograman [c. Bahasa pemrograman Javascript]	...	3. Pemahaman dan Penerapan Konsep [a. Pemahaman konsep XHTML]	3. Pemahaman dan Penerapan Konsep [b. Pemahaman dasar dan konsep Desain Grafis]	3. Pemahaman dan Penerapan Konsep [c. Pemahaman konsep Monitor dan administrasi Network Security]
8/10/2021 16:20:13	windaoktapias@student.untan.ac.id	Windari Oktapia Simanjuntak	D1041161032	Tidak Bisa	Tidak Bisa	Tidak Bisa	...	Tidak Bisa	Bisa	Tidak Bisa
8/10/2021 16:21:17	bravobb@student.untan.ac.id	Ansfridus Bravo	D1041161024	Tidak Bisa	Tidak Bisa	Tidak Bisa	...	Tidak Bisa	Bisa	Tidak Bisa
8/10/2021 16:23:11	rama.ulgasesa@student.untan.ac.id	Rama Ulgasesa	D1041161026	Tidak Bisa	Tidak Bisa	Tidak Bisa	...	Tidak Bisa	Bisa	Tidak Bisa
8/10/2021 16:26:20	farhan.hersinanda@student.untan.ac.id	Muhammad Farhan Hersinanda	D1041181026	Tidak Bisa	Tidak Bisa	Tidak Bisa	...	Tidak Bisa	Bisa	Tidak Bisa
8/10/2021 16:31:58	adityawijaya@student.untan.ac.id	Aditya Wijaya	D1041181038	Tidak Bisa	Bisa	Bisa	...	Tidak Bisa	Bisa	Bisa

3.3.3 Pengolahan Data

Data set yang digunakan sebagai data masukan adalah data yang didapatkan dari hasil kuesioner berupa kemampuan yang mahasiswa kuasai terhadap kriteria sub kemampuan yang dimiliki di setiap kelompok keahlian. Data set tersebut merupakan data yang belum dapat di proses klasifikasi, hal ini dikarenakan data belum memenuhi standar untuk sebuah *machine learning*, untuk itu diperlukan proses pengolahan data awal yaitu pelabelan dan inisialisasi.

3.3.3.1 Pelabelan

Pelabelan data adalah proses mengidentifikasi data mentah (gambar, file teks, video, dll.) dan menambahkan satu atau lebih label yang bermakna dan informatif untuk memberikan konteks sehingga model pembelajaran mesin dapat belajar darinya. Misalnya, label mungkin menunjukkan apakah foto berisi burung atau mobil, kata-kata apa yang diucapkan dalam rekaman audio, atau jika sinar-x mengandung tumor. Pelabelan data diperlukan untuk berbagai kasus penggunaan termasuk visi komputer, pemrosesan bahasa alami, dan pengenalan suara. Pelabelan data pada penelitian ini dilakukan berdasarkan bidang keahlian yang ada di Jurusan Informatika Fakultas Teknik Universitas Tanjungpura yaitu:

1. Kelompok Keahlian Sistem Informasi dan Data Spasial
2. Kelompok Keahlian Rekayasa Perangkat Lunak dan Komputasi Bergerak
3. Kelompok Keahlian Jaringan dan Keamanan
4. Kelompok Keahlian Komputasi dan Kecerdasan Buatan

Pelabelan mahasiswa ke dalam bidang keahlian ditentukan berdasarkan hasil perhitungan banyaknya sub kemampuan yang dikuasai oleh mahasiswa terhadap kriteria sub kemampuan yang sesuai dengan data sub kemampuan yang dimiliki di setiap kelompok keahlian. Adapun hasil dari pelabelan dapat dilihat pada tabel 3.3.

Tabel 3. 3 Pelabelan

Nama	Nim	1. Penguasaan Bahasa Pemrograman [a. Bahasa pemrograman XML]	1. Penguasaan Bahasa Pemrograman [b. Bahasa pemrograman Java]	1. Penguasaan Bahasa Pemrograman [c. Bahasa pemrograman Javascript]	...	3. Pemahaman dan Penerapan Konsep [a. Pemahaman konsep XHTML]	3. Pemahaman dan Penerapan Konsep [b. Pemahaman dasar dan konsep Desain Grafis]	3. Pemahaman dan Penerapan Konsep [c. Pemahaman konsep Monitor dan administrasi Network Security]	Label 1
Windari Oktapia Simanjuntak	D1041161032	Tidak Bisa	Tidak Bisa	Tidak Bisa	...	Tidak Bisa	Bisa	Tidak Bisa	Kelompok Keahlian Komputasi dan Kecerdasan Buatan
Ansfridus Bravo	D1041161024	Tidak Bisa	Tidak Bisa	Tidak Bisa	...	Tidak Bisa	Bisa	Tidak Bisa	Kelompok Keahlian Rekayasa Perangkat Lunak dan Komputasi Bergerak
Rama Ulgasesa	D1041161026	Tidak Bisa	Tidak Bisa	Tidak Bisa	...	Tidak Bisa	Bisa	Tidak Bisa	Kelompok Keahlian Komputasi dan Kecerdasan Buatan
Muhammad Farhan Hersinanda	D1041181026	Tidak Bisa	Tidak Bisa	Tidak Bisa	...	Tidak Bisa	Bisa	Tidak Bisa	Kelompok Keahlian Sistem Informasi dan Data Spasial
Aditya Wijaya	D1041181038	Tidak Bisa	Bisa	Bisa	...	Tidak Bisa	Bisa	Bisa	Kelompok Keahlian Jaringan dan Keamanan

3.3.3.2 Inisialisasi

Tools yang digunakan sebagai alat bantu untuk mengolah data dalam penelitian ini adalah *Jupyter Notebook* dengan menggunakan bahasa pemrograman Python. Data yang digunakan pada penelitian ini adalah data tingkat kemampuan yang dikuasai oleh mahasiswa terhadap sub kemampuan yang dimiliki di setiap kelompok keahlian Jurusan Informatika Fakultas Teknik Universitas Tanjungpura yang sudah dilabel pada tahapan pelabelan data. Tahapan yang dilakukan adalah menginisialisasi data ke variabel “DF” dalam bentuk dataframe dengan cara membaca “Dataset.csv”. Proses inisialisasi dilakukan dengan menjalankan kode program 3.1, dan hasil inisialisasi dapat dilihat pada gambar 3.2.

Kode program 3. 1 Inisialisasi

```
import seaborn as sns
import matplotlib.pyplot as plt
import pandas as pd
import numpy as np
from sklearn.metrics import confusion_matrix
from sklearn.naive_bayes import BernoulliNB
from sklearn.metrics import classification_report
from sklearn.model_selection import cross_val_predict
```

	Timestamp	Email Address	Nama	Nim	1. Penguasaan Bahasa Pemrograman [a. Bahasa pemrograman XML]	1. Penguasaan Bahasa Pemrograman [b. Bahasa perograman Java]	1. Penguasaan Bahasa Pemrograman [c. Bahasa pemrograman Javascript]	1. Penguasaan Bahasa Pemrograman [d. Bahasa pemrograman PHP dan ASP]	1. Penguasaan Bahasa Pemrograman [e. Bahasa pemrograman HTML]
0	10/8/2021 16:20	windaoktapias@student.untan.ac.id	Windari Oktapia Simanjuntak	D1041161032	Tidak Bisa	Tidak Bisa	Tidak Bisa	Bisa	Bisa
1	10/8/2021 16:21	bravobb@student.untan.ac.id	Ansfridus Bravo	D1041161024	Tidak Bisa	Tidak Bisa	Tidak Bisa	Bisa	Bisa
2	10/8/2021 16:23	rama.ulgasesa@student.untan.ac.id	Rama Ulgasesa	D1041161026	Tidak Bisa	Tidak Bisa	Tidak Bisa	Bisa	Tidak Bisa
3	10/8/2021 16:26	farhan.hersinanda@student.untan.ac.id	Muhammad Farhan Hersinanda	D1041181026	Tidak Bisa	Tidak Bisa	Tidak Bisa	Bisa	Bisa
4	10/8/2021 16:31	adityawijaya@student.untan.ac.id	Aditya Wijaya	D1041181038	Tidak Bisa	Bisa	Bisa	Bisa	Bisa

Gambar 3. 2 Hasil inisialisasi

3.3.4 Preprocessing

Preprocessing dilakukan dengan tujuan untuk mentransformasikan data mentah, setelah itu baru didapatkan dari proses pengambilan data ke format data yang lebih efisien dalam proses klasifikasi. Tahapan *preprocessing* dilakukan dalam lima tahap yaitu *cleaning*, *ordinal encoding*, *data selection*, *smote* dan *feature selection chi square*. Data set yang telah melalui tahapan *preprocessing* berjumlah 128 data set, kemudian disimpan menjadi data baru dalam format .csv untuk digunakan dalam proses pemodelan. Proses penyimpanan hasil *preprocessing* dilakukan dengan menjalankan kode program 3.2.

Kode program 3. 2 Menyimpan Hasil *Preprocessing*

```
df2=X
df2['label1']=y
df2
df2.to_csv("DatasetPreprocessing.csv")
```

3.3.4.1 *Cleaning*

Cleaning adalah salah satu tahapan dalam *preprocessing* yaitu proses analisa mengenai kualitas dari data, bisa juga mengoreksi ataupun menghapus data tersebut untuk disiapkan sebagai data yang akan digunakan dalam pembuatan model. Data yang di *cleaning* adalah data yang tidak struktur meliputi salah format dalam pengetikan, tidak akurat terlebih dahulu harus diperbaiki supaya data berkualitas diantaranya valid, akurat, komplit, konsisten dan seragam. Pada tahap ini, *cleaning* yang dilakukan adalah penulisan “Label 1” dan “Label 2” diubah menjadi “label1”, dan “label2” tanpa penggunaan huruf kapital dan spasi, memberikan *missing values* dengan nilai NA/NAN lalu menuliskan dengan tulisan “kosong”. Proses *cleaning* dilakukan dengan menjalankan kode program 3.3, dan hasil *cleaning* dapat dilihat pada gambar 3.3.

Kode program 3. 3 *Cleaning*

```
#Cleaning
df.rename(
    columns={ 'Label 2': 'label2', 'Label 1': 'label1' },
    inplace=True,
)
df = df.fillna(value="kosong")
df
```

1. Penguasaan Bahasa Pemrograman [f. Bahasa pemrograman SQL]	...	3. Pemahaman dan Penerapan Konsep [h. Pemahaman perancangan ERD dan UML]	3. Pemahaman dan Penerapan Konsep [i. Pemahaman perancangan arsitektur DBMS]	3. Pemahaman dan Penerapan Konsep [j. Pemahaman arsitektur RISC]	3. Pemahaman dan Penerapan Konsep [k. Pemahaman konsep manajemen perangkat lunak I/O]	3. Pemahaman dan Penerapan Konsep [l. Pemahaman teknik pengujian Model/Aplikasi]	3. Pemahaman dan Penerapan Konsep [m. Pemahaman salah satu atau lebih pembelajaran mesin atau kecerdasan buatan]	3. Pemahaman dan Penerapan Konsep [n. Pemahaman konsep mobile computing]	3. Pemahaman dan Penerapan Konsep [o. Memahami dan menguasai Search Engine Optimization (SEO)]	label1	label2
Tidak Bisa	...	Tidak Bisa	Tidak Bisa	Tidak Bisa	Tidak Bisa	Tidak Bisa	Bisa	Tidak Bisa	Tidak Bisa	Kelompok Keahlian Komputasi dan Kecerdasan Buatan	kosong
Bisa	...	Bisa	Bisa	Tidak Bisa	Bisa	Bisa	Bisa	Tidak Bisa	Tidak Bisa	Kelompok Keahlian Rekayasa Perangkat Lunak dan...	kosong
Tidak Bisa	...	Bisa	Bisa	Tidak Bisa	Tidak Bisa	Bisa	Bisa	Tidak Bisa	Tidak Bisa	Kelompok Keahlian Komputasi dan Kecerdasan Buatan	kosong
Bisa	...	Bisa	Bisa	Tidak Bisa	Bisa	Bisa	Bisa	Bisa	Tidak Bisa	Kelompok Keahlian Sistem Informasi dan Data Sp...	kosong
Bisa	...	Bisa	Bisa	Bisa	Tidak Bisa	Tidak Bisa	Tidak Bisa	Bisa	Tidak Bisa	Kelompok Keahlian Jaringan dan Keamanan	kosong

Gambar 3. 3 Hasil *cleaning*

3.3.4.2 Ordinal Encoding

Data memiliki fitur nama yang menunjukkan nama mahasiswa Informatika, dan fitur sub kemampuan yang menunjukkan tingkat kemampuan yang mahasiswa kuasai pada kelompok keahlian, data tersebut merupakan data kategorikal dan terlebih dahulu diubah menjadi data *numerical*. Proses ini disebut pengkodean ordinal atau *Ordinal Encoding*. Transformasi pengkodean ordinal ini tersedia di perpustakaan pembelajaran mesin Python *scikit-learn* melalui kelas *OrdinalEncoder*. Pada penelitian ini kategori yang digunakan terdapat di tingkat kemampuan yang dikuasai mahasiswa pada sub kemampuan dan label, kategori tingkat kemampuan pada sub kemampuan adalah “Tidak Bisa” dan “Bisa” setiap kategori diberi nilai integer dengan mengubah “Tidak Bisa” dan “Bisa” menjadi bilangan bulat yaitu Tidak bisa : 0, dan Bisa : 1, secara *default* ini juga diterapkan pada kategori label dalam urutan 0 sampai 3, dimana kategori label adalah “Kelompok Keahlian Sistem Informasi dan Data Spasial : 0”, “Kelompok Keahlian Rekayasa Perangkat Lunak dan Komputasi Bergerak : 1”, “Kelompok Keahlian Jaringan dan Keamanan : 2”, “Kelompok Keahlian Komputasi dan Kecerdasan Buatan : 3”. Proses *ordinal encoding* dilakukan dengan menjalankan kode program 3.4, dan hasil *ordinal encoding* dapat dilihat pada gambar 3.4.

Kode program 3. 4 Ordinal Encoding

```
#Ordinal Encoding

df = df.replace("Tidak Bisa", 0)
df = df.replace("Bisa", 1)
df.label1 = df.label1.replace("Kelompok Keahlian Sistem Informasi dan Data Spasial", 0)
df.label1 = df.label1.replace("Kelompok Keahlian Rekayasa Perangkat Lunak dan Komputasi Bergerak", 1)
df.label1 = df.label1.replace("Kelompok Keahlian Jaringan dan Keamanan", 2)
df.label1 = df.label1.replace("Kelompok Keahlian Komputasi dan Kecerdasan Buatan", 3)
df.label2 = df.label2.replace("Kelompok Keahlian Sistem Informasi dan Data Spasial", 0)
df.label2 = df.label2.replace("Kelompok Keahlian Rekayasa Perangkat Lunak dan Komputasi Bergerak", 1)
df.label2 = df.label2.replace("Kelompok Keahlian Jaringan dan Keamanan", 2)
df.label2 = df.label2.replace("Kelompok Keahlian Komputasi dan Kecerdasan Buatan", 3)
df
```

1. Penguasaan Bahasa Pemrograman [e. Bahasa pemrograman HTML]	1. Penguasaan Bahasa Pemrograman [f. Bahasa pemrograman SQL]	...	3. Pemahaman dan Penerapan Konsep [h. Pemahaman perancangan ERD dan UML]	3. Pemahaman dan Penerapan Konsep [i. Pemahaman perancangan arsitektur DBMS]	3. Pemahaman dan Penerapan Konsep [j. Pemahaman arsitektur RISC]	3. Pemahaman dan Penerapan Konsep [k. Pemahaman konsep manajemen perangkat lunak I/O]	3. Pemahaman dan Penerapan Konsep [l. Pemahaman teknik pengujian Model/Aplikasi]	3. Pemahaman dan Penerapan Konsep [m. Pemahaman salah satu atau lebih pembelajaran mesin atau kecerdasan buatan]	3. Pemahaman dan Penerapan Konsep [n. Pemahaman konsep mobile computing]	3. Pemahaman dan Penerapan Konsep [o. Memahami dan menguasai Search Engine Optimization (SEO)]	label1	label2
1	0	...	0	0	0	0	0	1	0	0	3	kosong
1	1	...	1	1	0	1	1	1	0	0	1	kosong
0	0	...	1	1	0	0	1	1	0	0	3	kosong
1	1	...	1	1	0	1	1	1	1	0	0	kosong
1	1	...	1	1	1	0	0	0	1	0	2	kosong
...
0	1	...	1	1	1	1	1	1	1	0	0	kosong
1	1	...	1	1	0	1	1	1	1	1	2	kosong
1	1	...	0	1	0	1	1	1	0	0	0	kosong
1	1	...	1	1	0	1	0	1	0	0	2	kosong
1	1	...	1	1	0	1	0	0	1	1	0	kosong

Gambar 3. 4 Hasil ordinal encoding

3.3.4.3 Data Selection

Data Selection merupakan proses meminimalkan jumlah data yang akan digunakan sebagai data set, maka dilakukan tahapan seleksi data terlebih dahulu dengan tetap merepresentasikan data aslinya. Data yang diseleksi adalah data hasil kuesioner dimana terdapat mahasiswa yang menunjukkan hasil dari pelabelan memiliki lebih dari 1 kelompok keahlian. Proses *data selection* dilakukan dengan menjalankan kode program 3.5, dan hasil *data selection* dapat dilihat pada gambar 3.5.

Kode program 3. 5 Data Selection

```
#Data Selection
#Delete data jika terdapat kelompok keahlian pada label2
df = df[df["label2"] == "kosong"]
# df.drop(columns=['label2'])

#Data Label/ Data Dependent
y = df.label1

#Data Independent
features = df.columns[4:61]
X = df[features]

#Select Feature if datatype = integer
discrete_features = X.dtypes == int
X
```

1. Penguasaan Bahasa Pemrograman [j. Bahasa pemrograman Curl]	3. Pemahaman dan Penerapan Konsep [g. Pemahaman konsep Database]	3. Pemahaman dan Penerapan Konsep [h. Pemahaman perancangan ERD dan UML]	3. Pemahaman dan Penerapan Konsep [i. Pemahaman perancangan arsitektur DBMS]	3. Pemahaman dan Penerapan Konsep [j. Pemahaman arsitektur RISC]	3. Pemahaman dan Penerapan Konsep [k. Pemahaman konsep manajemen perangkat lunak I/O]	3. Pemahaman dan Penerapan Konsep [l. Pemahaman teknik pengujian Model/Aplikasi]	3. Pemahaman dan Penerapan Konsep [m. Pemahaman salah satu atau lebih pembelajaran mesin atau kecerdasan buatan]	3. Pemahaman dan Penerapan Konsep [n. Pemahaman konsep mobile computing]	3. Pemahaman dan Penerapan Konsep [o. Memahami dan menguasai Search Engine Optimization (SEO)]	label1
0 ...	0	0	0	0	0	0	1	0	0	3
0 ...	1	1	1	0	1	1	1	0	0	1
0 ...	0	1	1	0	0	1	1	0	0	3
0 ...	1	1	1	0	1	1	1	1	0	0
0 ...	1	1	1	1	0	0	0	1	0	2
...
1 ...	1	1	1	1	1	1	1	1	0	0
0 ...	1	1	1	0	1	1	1	1	1	2
0 ...	1	0	1	0	1	1	1	0	0	0
1 ...	1	1	1	0	1	0	1	0	0	2
0 ...	1	1	1	0	1	0	0	1	1	0

Gambar 3. 5 Hasil data selection

3.3.4.4 SMOTE

Teknik ini menyintesis sampel baru dari kelas minoritas untuk menyeimbangkan data set dengan cara membuat *instance* baru dari *minority class* dengan pembentukan *convex* kombinasi dari *instance* yang berdekatan. Dengan metode ini dapat membuat data set menjadi seimbang tanpa terlalu *overfit*, yaitu dengan membuat *sample synthetic* daripada dengan melakukan *duplicate sample*. Total data set sebelum dilakukan tahapan *smote* berjumlah 128 data yang terdiri dari 57 pada kelas 0, 18 pada kelas 1, 46 pada kelas 2 dan 7 pada kelas 3 (lihat gambar 3.6). Pada tahap ini penyeimbangan dilakukan pada kelas 1, 2 dan 3, setelah dilakukan tahapan *smote*, kelas 1, 2 dan 3 menjadi seimbang dengan total keseluruhan data set berjumlah 228 data yang terdiri dari 57 pada kelas 0, 57 pada

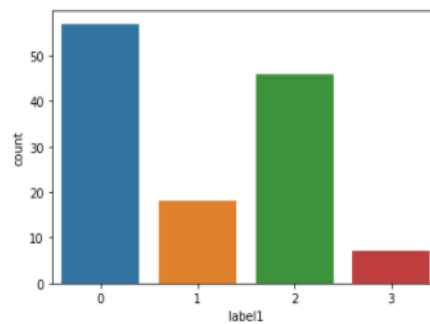
kelas 1, 57 pada kelas 2 dan 57 pada kelas 3 (lihat gambar 3.7). Proses *smote* dilakukan dengan menjalankan kode program 3.6, dan hasil *smote* dapat dilihat pada gambar 3.8.

Kode program 3. 6 SMOTE

```
#Importing SMOTE
#from imblearn import under_sampling, over_sampling
from imblearn.over_sampling import SMOTE
#Oversampling the data

smote = SMOTE()
Xsmt, ysmt = smote.fit_resample(X, y)

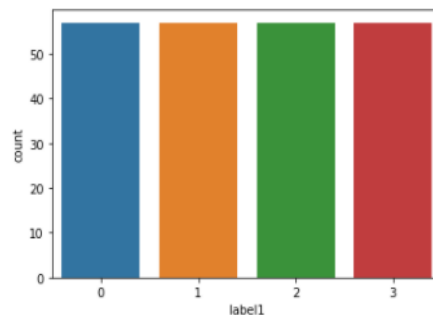
#Creating a new Oversampling Data Frame
#df_oversampler = pd.DataFrame(X, columns = features)
#df_oversampler['label1']
sns.countplot(df['label1'])
```



```
df.label1.value_counts()
```

```
0    57
2    46
1    18
3     7
Name: label1, dtype: int64
```

Gambar 3. 6 Data sebelum *smote*



```
df.label1.value_counts()
```

```
0    57
1    57
2    57
3    57
Name: label1, dtype: int64
```

Gambar 3. 7 Data sesudah *smote*


```
df = pd.read_csv("DatasetSmote.csv", index_col="Unnamed: 0")
df
```

	1. Penguasaan Bahasa Pemrograman [a. Bahasa pemrograman XML]	1. Penguasaan Bahasa Pemrograman [b. Bahasa pemrograman Java]	1. Penguasaan Bahasa Pemrograman [c. Bahasa pemrograman Javascript]	1. Penguasaan Bahasa Pemrograman [d. Bahasa pemrograman PHP dan ASP]	1. Penguasaan Bahasa Pemrograman [e. Bahasa pemrograman HTML]	1. Penguasaan Bahasa Pemrograman [f. Bahasa pemrograman SQL]	1. Penguasaan Bahasa Pemrograman [g. Bahasa pemrograman Prolog]	1. Penguasaan Bahasa Pemrograman [h. Bahasa pemrograman RUBY]	1. Penguasaan Bahasa Pemrograman [i. Bahasa pemrograman Python]	1. Penguasaan Bahasa Pemrograman [j. Bahasa pemrograman Curl]	...
0	0	0	0	1	1	0	0	0	1	0	...
1	0	0	0	1	1	1	0	0	1	0	...
2	0	0	0	1	0	0	0	0	1	0	...
3	0	0	0	1	1	1	0	0	1	0	...
4	0	1	1	1	1	1	0	0	0	0	...
...
223	0	0	0	0	0	0	0	0	1	0	...
224	0	0	0	1	0	0	0	0	1	0	...
225	0	0	0	1	0	0	0	0	1	0	...
226	0	0	0	1	1	0	0	0	1	0	...
227	0	0	0	1	0	0	0	0	1	0	...

228 rows × 57 columns

Gambar 3. 8 Hasil *smote*

Data hasil *smote* disimpan menjadi data set baru ke dalam format .csv. Proses penyimpanan hasil *smote* dilakukan dengan menjalankan kode program 3.7.

Kode program 3. 7 Menyimpan hasil *smote*

```
df2=Xsmt
df2['label1']=ysmt
df2
df2.to_csv("DatasetSmote.csv")
```

3.3.4.5 Feature Selection

Seleksi fitur (*feature selection*) dilakukan untuk mereduksi fitur sampel yang tidak relevan dalam proses klasifikasi oleh *Naive Bayes*, seleksi fitur yang digunakan pada penelitian ini yaitu dengan menggunakan *Chi Square*. Seleksi fitur *Chi Square* menggunakan teori statistika untuk menguji independensi sebuah term dengan kategorinya. Salah satu tujuan penggunaan seleksi fitur adalah untuk menghilangkan fitur yang tidak memiliki korelasi dalam klasifikasi. Dalam seleksi fitur *Chi Square* berdasarkan teori statistika, dua peristiwa di antaranya adalah, kemunculan dari fitur dan kemunculan dari kategori, yang kemudian setiap nilai term diurutkan dari yang tertinggi. Uji *Chi Square* dalam statistika diterapkan untuk menguji independensi dari dua peristiwa. Proses *Chi Square* dilakukan dengan menjalankan kode program 3.8, dan hasil *Chi Square* dapat dilihat pada gambar 3.9.

Kode program 3. 8 *Chi Square*

```
#Module Chi-square
def ChiSquare_Test(df, ycol, xcol, alpha):

    result = {}

    #Looping for every x variable in Dataframe
    for col in xcol :
        crosstab = pd.crosstab(df[ycol],df[col])
        _, p, _, _ = chi2_contingency(crosstab)

        #logic
        if p <= alpha :
            result[col] = 'Correlated'
        else :
            result[col] = 'Not Correlated'

    #The result will be created to the pandas dataframe
    df = pd.DataFrame.from_dict(result, orient='Index', columns=['Decision'])
    df.index.name='Colname'

    return df

def list_dtypes(df):
    categorical_list = []
    numerical_list = []
    for col in df.columns.tolist():
        if df[col].dtype=='object':
            categorical_list.append(col)
        else:
            numerical_list.append(col)

    return categorical_list, numerical_list
```

```
## seperating data types
categorical_list, numerical_list = list_dtypes(df1)
categorical_list = categorical_list[4:60]
```

```
## Check correlation between categorical independent and categorical dependent (Chisquare Test)
# 1. Define Dataframe
categorical_variable_list = categorical_list + ['label1']
df_categ = df[categorical_variable_list]
df_categ['label1'] = df_categ['label1'].astype(object)

# 2. Input y column
y = 'label1'

# 3. Input X columns
features = df.drop(columns=['label1'])
x = features

# 4. Significance Level
a = 0.05

# 5. Chi-Square Analysis
categorical_test = ChiSquare_Test(df = df_categ, ycol=y, xcol= x, alpha= a)
categorical_test
```

```
[ '1. Penguasaan Bahasa Pemrograman [a. Bahasa pemrograman XML]',
'1. Penguasaan Bahasa Pemrograman [b. Bahasa pemrograman Java]',
'1. Penguasaan Bahasa Pemrograman [c. Bahasa pemrograman Javascript]',
'1. Penguasaan Bahasa Pemrograman [e. Bahasa pemrograman HTML]',
'1. Penguasaan Bahasa Pemrograman [f. Bahasa pemrograman SQL]',
'1. Penguasaan Bahasa Pemrograman [i. Bahasa pemrograman Python]',
'1. Penguasaan Bahasa Pemrograman [n. Bahasa pemrograman Visual Basic]',
'1. Penguasaan Bahasa Pemrograman [o. Bahasa pemrograman C atau C++]',
'2. Penguasaan Tools/Aplikasi [a. Tools Apache Spark]',
'2. Penguasaan Tools/Aplikasi [b. Tools Matplotlib]',
'2. Penguasaan Tools/Aplikasi [c. Tools NLTK]',
'2. Penguasaan Tools/Aplikasi [e. Tools Jupiter]',
'2. Penguasaan Tools/Aplikasi [f. Tools Moses]',
'2. Penguasaan Tools/Aplikasi [h. Tools Numpy]',
'2. Penguasaan Tools/Aplikasi [i. Tools Scipy]',
'2. Penguasaan Tools/Aplikasi [j. Tools Scikit-Learn]',
'2. Penguasaan Tools/Aplikasi [k. Tools TensorFlow]',
'2. Penguasaan Tools/Aplikasi [o. NET cut, Cain dan Abel, Wireshark]',
'2. Penguasaan Tools/Aplikasi [p. Ms SQL]',
'2. Penguasaan Tools/Aplikasi [q. Ms Project]',
'2. Penguasaan Tools/Aplikasi [r. StarUML]',
'2. Penguasaan Tools/Aplikasi [s. Packet Tracer]',
'2. Penguasaan Tools/Aplikasi [t. Winbox]',
'2. Penguasaan Tools/Aplikasi [u. Netbeans]',
'2. Penguasaan Tools/Aplikasi [v. Eclipse]',
'2. Penguasaan Tools/Aplikasi [z. SQLite]',
'3. Pemahaman dan Penerapan Konsep [a. Pemahaman konsep XHTML]',
'3. Pemahaman dan Penerapan Konsep [b. Pemahaman dasar dan konsep Desain Grafis]',
'3. Pemahaman dan Penerapan Konsep [c. Pemahaman konsep Monitor dan administrasi Network Security]',
'3. Pemahaman dan Penerapan Konsep [d. Penerapan konsep FTP server dan Web server]',
'3. Pemahaman dan Penerapan Konsep [e. Pemahaman Basic IP dan Subnetting]',
'3. Pemahaman dan Penerapan Konsep [f. Pemahaman sistem informasi dan jenis sistem informasi : MIS, DSS, dan EIS]',
'3. Pemahaman dan Penerapan Konsep [g. Pemahaman konsep Database]',
'3. Pemahaman dan Penerapan Konsep [i. Pemahaman perancangan arsitektur DBMS]',
'3. Pemahaman dan Penerapan Konsep [j. Pemahaman arsitektur RISC]',
'3. Pemahaman dan Penerapan Konsep [k. Pemahaman konsep manajemen perangkat lunak I/O]',
'3. Pemahaman dan Penerapan Konsep [l. Pemahaman teknik pengujian Model/Aplikasi]',
'3. Pemahaman dan Penerapan Konsep [m. Pemahaman salah satu atau lebih pembelajaran mesin atau kecerdasan buatan]',
'3. Pemahaman dan Penerapan Konsep [n. Pemahaman konsep mobile computing]',
'3. Pemahaman dan Penerapan Konsep [o. Memahami dan menguasai Search Engine Optimization (SEO)]'
```

Gambar 3. 9 Hasil *chi square*

3.3.5 Pemodelan

Pembuatan model berdasarkan kesesuaian data dengan metode yang paling baik dari beberapa metode pengklasifikasian yang sudah digunakan oleh beberapa peneliti sebelumnya yaitu *Bernoulli Naive Bayes* yang merupakan salah satu jenis algoritma *Naive Bayes*. Pemodelan dengan menggunakan algoritma *Bernoulli Naive Bayes* dikarenakan data yang digunakan adalah data *numerical*. Tahapan untuk menjalani proses *Bernoulli Naive Bayes* sebagai berikut:

1. Menghitung jumlah kelas / label “P(H)”
2. Menghitung jumlah kasus per kelas “P(X|H)”
3. Kalikan semua variabel kelas “ $P(X|H) \times P(H)$ ”
4. Bandingkan hasil per kelas

Perhitungan model klasifikasi dilakukan dengan menggunakan algoritma *Bernoulli Naive Bayes* ditunjukkan pada persamaan (3.1).

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \quad (3.1)$$

Jika ada satu data yang memiliki nilai 0, maka dilakukan *Laplacian Correction* untuk menambahkan nilai 1 (satu) yang ada pada atribut/kriteria, ditunjukkan pada persamaan (3.2).

$$P(t | c) = \frac{P(Tct) + \text{Smoothing}}{(\sum Tct)_{t \in V} + |V|} \quad (3.2)$$

Pada tahap ini pemodelan dilakukan dengan memanggil fungsi *Bernoulli Naive Bayes* dari *library Scikit Learn* untuk melakukan proses training pada data yang telah melalui tahapan *preprocessing* dan *feature selection*. Proses pemodelan dilakukan dengan menjalankan kode program 3.9.

Kode program 3. 9 *BernoulliNB*

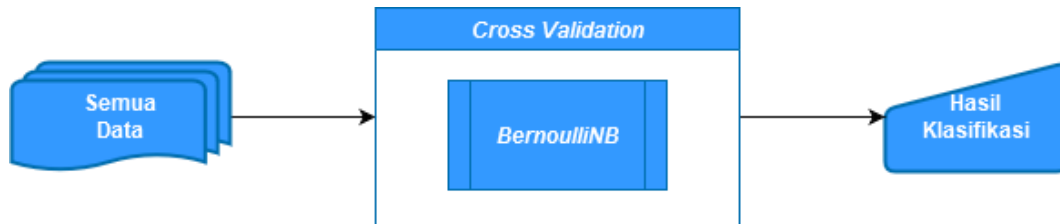
```
# Mengaktifkan/memanggil/membuat fungsi klasifikasi Naive Bayes
modelnb = BernoulliNB()
# Memasukkan data training pada fungsi klasifikasi Naive Bayes
nbtrain = modelnb.fit(Xchi, y)
```

3.3.6 Validation

Pembagian data ke dalam data latih dan data uji juga diperlukan untuk pembuatan model, tahap training dan testing akan divalidasi menggunakan *cross validation* dengan 10 kali validasi. Hasil dari tahap ini adalah nilai *precision*, *recall*, dan *accuracy*, nilai ini akan dibandingkan untuk mengetahui model manakah yang paling baik, selanjutnya model yang sudah dibuat akan disimpan dan digunakan untuk melakukan prediksi kelompok keahlian mahasiswa. Proses *cross validation* dilakukan dengan menjalankan kode program 3.10, dan tahapan *cross validation* dapat dilihat pada gambar 3.10.

Kode program 3. 10 K-Fold

```
#Cross Validation
predict = cross_val_predict(estimator = modelnb, X = Xchi, y = y, cv = 10)
```



Gambar 3. 10 Tahapan *cross validation*

3.3.7 Pengujian

Pengujian dilakukan dengan menggunakan empat skenario pengujian. Skenario pengujian pertama dilakukan menggunakan algoritma *Naive Bayes*, skenario pengujian kedua dilakukan menggunakan algoritma *Naive Bayes* dan *feature selection Chi Square*, skenario pengujian ketiga dilakukan menggunakan algoritma *Naive Bayes* dan *SMOTE*, skenario pengujian keempat dilakukan menggunakan algoritma *Naive Bayes*, *SMOTE* dan *feature selection Chi Square*. Pada skenario pengujian pertama dan kedua dilakukan menggunakan data yang sudah melalui tahapan *preprocessing cleaning*, *ordinal encoding*, dan *data selection* yang berjumlah **128** data set. Sedangkan untuk skenario pengujian ketiga dan keempat, dilakukan menggunakan data yang sudah melalui tambahan tahapan pada *preprocessing* yaitu dengan menggunakan *SMOTE* yang berjumlah **228** data set.

Tabel 3. 4 Skenario Pengujian

Skenario Pengujian	<i>SMOTE</i>	<i>Feature Selection Chi Square</i>
Skenario Pengujian Pertama	×	×
Skenario Pengujian Kedua	×	✓
Skenario Pengujian Ketiga	✓	×
Skenario Pengujian Keempat	✓	✓

Untuk menguji performa klasifikasi yang dihasilkan, dibutuhkan metode pengujian, adapun pengujian yang dilakukan adalah pengujian dengan menggunakan pengujian *Confusion Matrix*. Pengujian *Confusion Matrix* merupakan dimana model yang digunakan diuji dengan cara melakukan perhitungan akurasi yang bertujuan untuk menyatakan jumlah data uji yang benar diklasifikasikan dan jumlah data uji yang salah diklasifikasikan terhadap kelas yang seharusnya serta untuk mengetahui kemungkinan mahasiswa masuk ke dalam kelas (kelompok keahlian) tertentu. Akurasi adalah tingkat kedekatan antara nilai prediksi dan nilai aktual (Han *et al.*, 2012). Maka akurasi adalah parameter yang sangat penting dalam pengujian, karena mempresentasikan seberapa besar keberhasilan mesin dalam melakukan klasifikasi. Proses *confusion matrix* dilakukan dengan menjalankan kode program 3.11, dan hasil *confusion matrix* dapat dilihat pada Lampiran F.

Kode program 3. 11 *Confusion Matrix*

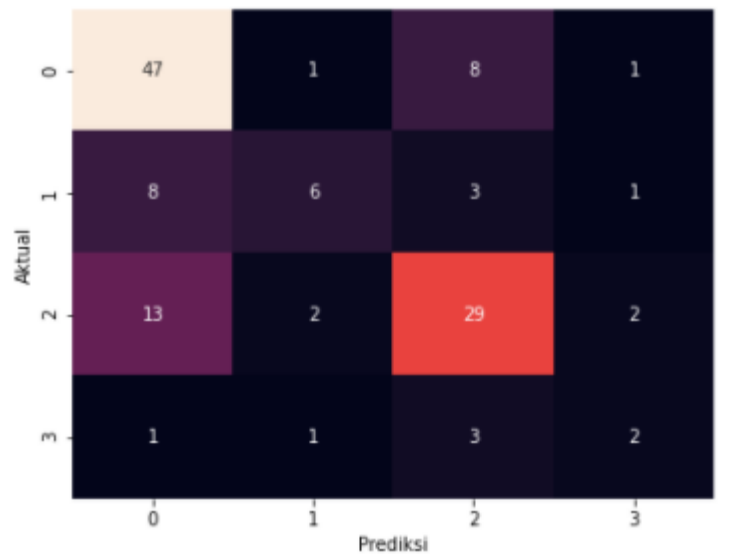
```
print("Confusion Matrix: \n",confusion_matrix(y, predict), "\n")
print("Classification Report: \n",classification_report(y, predict))
```

BAB IV HASIL DAN ANALISIS

4.1 Hasil Pengujian

Berikut hasil dari pengujian Klasifikasi Bidang Keahlian Mahasiswa Menggunakan Algoritma *Naive Bayes*. Adapun hasil skenario pengujian yang dilakukan adalah sebagai berikut:

1. Skenario pengujian pertama dilakukan menggunakan algoritma *Naive Bayes*.

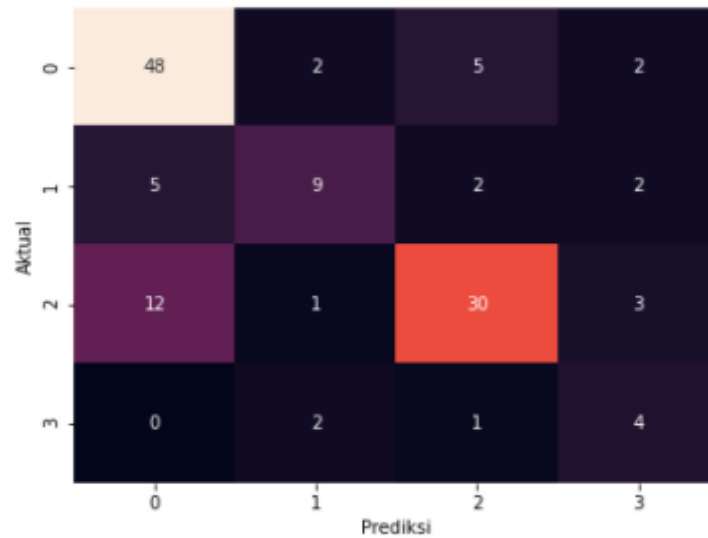


Gambar 4. 1 Hasil skenario pengujian 1

Pengujian ini menggunakan data set hasil *preprocessing* yang berjumlah 128 data, perhitungan *accuracy* dilakukan dengan rumus berikut:

$$\begin{aligned} Accuracy &= \frac{TP+TN}{TP+TN+FP+FN} \\ &= \frac{47+37}{47+37+28+16} \\ &= 0.66 = 66\% \end{aligned} \tag{4.1}$$

2. Skenario pengujian kedua dilakukan menggunakan algoritma *Naive Bayes* dan *feature selection Chi Square*.

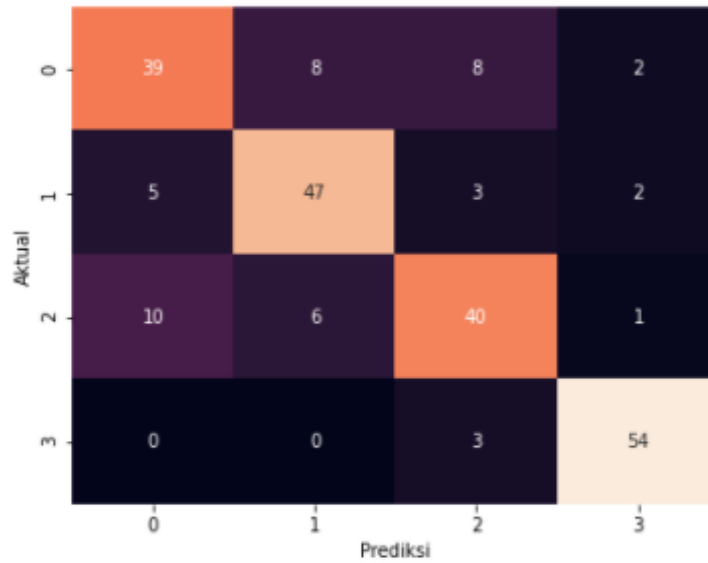


Gambar 4. 2 Hasil skenario pengujian 2

Pengujian ini menggunakan data set hasil *preprocessing* yang berjumlah 128 data, perhitungan *accuracy* dilakukan dengan rumus berikut:

$$\begin{aligned}
 Accuracy &= \frac{TP+TN}{TP+TN+FP+FN} & (4.2) \\
 &= \frac{9+82}{9+82+21+16} \\
 &= 0.71 = 71\%
 \end{aligned}$$

3. Skenario pengujian ketiga dilakukan menggunakan algoritma *Naive Bayes* dan *SMOTE*.

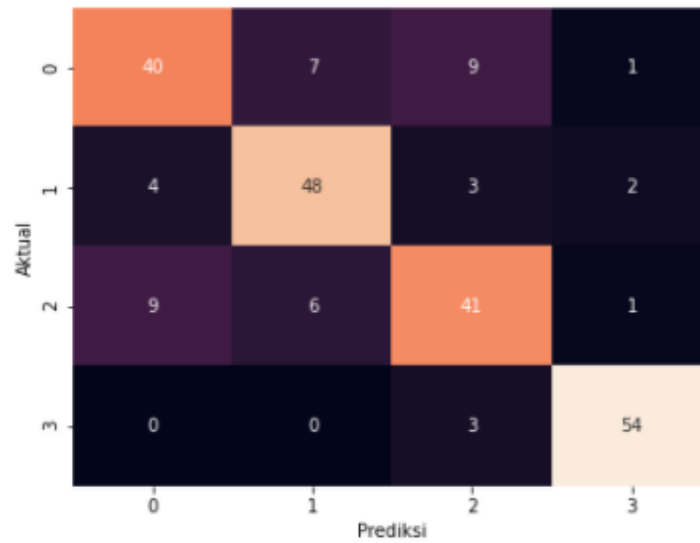


Gambar 4. 3 Hasil skenario pengujian 3

Pengujian ini menggunakan data set hasil *SMOTE* yang berjumlah 228 data, perhitungan *accuracy* dilakukan dengan rumus berikut:

$$\begin{aligned}
 Accuracy &= \frac{TP+TN}{TP+TN+FP+FN} & (4.3) \\
 &= \frac{40+140}{40+140+24+24} \\
 &= 0.79 = 79\%
 \end{aligned}$$

4. Skenario pengujian keempat dilakukan menggunakan algoritma *Naive Bayes*, *SMOTE* dan *feature selection Chi Square*.



Gambar 4. 4 Hasil skenario pengujian 4

Pengujian ini menggunakan data set hasil *SMOTE* yang berjumlah 228 data, perhitungan accuracy dilakukan dengan rumus berikut:

$$\begin{aligned}
 Accuracy &= \frac{TP+TN}{TP+TN+FP+FN} & (4.4) \\
 &= \frac{54+129}{54+129+22+23} \\
 &= 0.80 = 80\%
 \end{aligned}$$

Berikut merupakan hasil pengujian dari keempat skenario pengujian dijabarkan ke dalam Tabel 4.1.

Tabel 4. 1 Hasil Pengujian

Parameter pengujian	Skenario 1	Skenario 2	Skenario 3	Skenario 4
TP + TN	84	91	180	183
FN	16	16	24	23
FP	28	21	24	22
Accuracy	60%	71%	79%	80%

Keterangan:

1. Pengujian yang dilakukan pada skenario pengujian 1, 2, 3 dan 4 dengan parameter pengujian *accuracy*.
2. Penjumlahan TP + TN adalah jumlah testing data yang diprediksi dengan benar oleh model. Sedangkan hasil penjumlahan dari TP + TN + FP + FN adalah jumlah keseluruhan data yang digunakan dalam testing data (nilai aktual).

4.2 Analisis Hasil Pengujian

Analisis hasil pengujian Klasifikasi Bidang Keahlian Mahasiswa Menggunakan Algoritma *Naive Bayes* dengan parameter pengujian *accuracy* sebagai berikut:

Pengujian dengan menggunakan **Skenario pengujian 1**, dari **128** data set sebanyak **84** data berhasil diprediksi dengan benar oleh model, diantaranya **47** data pada kelompok keahlian Sistem Informasi dan Data Spasial (diinisialisasikan sebagai SI), **6** data pada kelompok keahlian Rekayasa Perangkat Lunak dan Komputasi Bergerak (diinisialisasikan sebagai RPL), **29** data pada kelompok keahlian Jaringan dan Keamanan (diinisialisasikan sebagai JK), dan **2** data pada kelompok keahlian Komputasi dan Kecerdasan Buatan (diinisialisasikan sebagai AI). Sedangkan sebanyak **44** data gagal diprediksi dengan benar oleh model. Hasil pengujian klasifikasi menunjukkan bahwa performa dari model *Naive Bayes* lemah terhadap data set yang mempunyai kelas data yang tidak seimbang dan fitur yang tidak relevan. Sehingga menghasilkan performa model dengan nilai *accuracy* sebesar **66%**.

Pengujian dengan menggunakan **Skenario pengujian 2**, dari **128** data set sebanyak **91** data berhasil diprediksi dengan benar oleh model, diantaranya **48** data pada kelompok keahlian SI, **9** data pada kelompok keahlian RPL, **30** data pada kelompok keahlian JK, dan **4** data pada kelompok keahlian AI. Sedangkan sebanyak **37** data gagal diprediksi dengan benar oleh model. Hasil pengujian klasifikasi pada model *Naive Bayes* dengan penambahan *feature selection Chi Square* memiliki penggunaan fitur yang berbeda. Secara umum *Chi Square* tidak terlalu mempengaruhi urutan hasil dari klasifikasi, tetapi mampu memberikan peningkatan pada bidang keahlian mahasiswa yang terprediksi dengan benar oleh model, dikarenakan data set yang digunakan sudah memiliki fitur yang relevan. Sehingga menghasilkan performa model dengan nilai *accuracy* sebesar **71%**.

Pengujian dengan menggunakan **Skenario pengujian 3**, dari **228** data set sebanyak **180** data berhasil diprediksi dengan benar oleh model, diantaranya **39** data pada kelompok keahlian SI, **47** data pada kelompok keahlian RPL, **40** data

pada kelompok keahlian JK, dan **54** data pada kelompok keahlian AI. Sedangkan sebanyak **48** data gagal diprediksi dengan benar oleh model. Hasil pengujian klasifikasi menunjukkan bahwa model *Naive Bayes* dengan penambahan *SMOTE* memberikan peningkatan pada bidang keahlian mahasiswa yang terprediksi dengan benar oleh model, dikarenakan data set yang digunakan sudah memiliki kelas data yang seimbang. Sehingga menghasilkan performa model dengan nilai *accuracy* sebesar **79%**.

Pengujian dengan menggunakan **Skenario pengujian 4**, dari **228** data set sebanyak **183** data berhasil diprediksi dengan benar oleh model, diantaranya **40** data pada kelompok keahlian SI, **48** data pada kelompok keahlian RPL, **41** data pada kelompok keahlian JK, dan **54** data pada kelompok keahlian AI. Sedangkan sebanyak **45** data gagal diprediksi dengan benar oleh model. Untuk hasil performa klasifikasi setiap model *Naive Bayes*, menunjukkan bahwa model *Naive Bayes* dengan penambahan *SMOTE* dan *feature selection Chi Square* lebih berpengaruh terhadap hasil klasifikasi. Sehingga menghasilkan performa model dengan nilai *accuracy* sebesar **80%**.

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Pada hasil dan analisa hasil pengujian Klasifikasi Bidang Keahlian Mahasiswa Menggunakan Algoritma *Naive Bayes*, maka dapat ditarik kesimpulan sebagai berikut:

1. Metode Klasifikasi dengan menggunakan algoritma *Naive Bayes* menghasilkan model yang dapat memprediksi bidang keahlian mahasiswa Informatika Fakultas Teknik Universitas Tanjungpura.
2. Penggunaan tingkat kemampuan yang dikuasai oleh mahasiswa terhadap kriteria sub kemampuan yang dimiliki di setiap kelompok keahlian sebagai data, dapat memprediksi bidang keahlian mahasiswa Informatika Fakultas Teknik Universitas Tanjungpura.
3. Penggunaan *SMOTE* lebih berpengaruh terhadap hasil pengujian klasifikasi pada model *Naive Bayes* yang dapat meningkatkan nilai *accuracy* sebesar 79%, dibandingkan dengan model *Naive Bayes* dengan penambahan *feature selection Chi Square* yang hanya dapat meningkatkan nilai *accuracy* sebesar 71%. Hal ini dikarenakan hasil skenario pengujian 2 menunjukkan performa dari model *Naive Bayes* dengan penambahan *feature selection Chi Square* masih belum berpengaruh terhadap data set yang mempunyai kelas data yang tidak seimbang. Sedangkan hasil skenario pengujian 3 menunjukkan bahwa terjadi kenaikan bidang keahlian mahasiswa yang terprediksi dengan benar oleh model pada kelas RPL, JK dan AI. Sehingga memberikan performa model dengan keakuratan lebih besar.
4. Pengujian klasifikasi pada model *Naive Bayes* jika menggunakan

SMOTE dan *feature selection Chi Square* memberikan performa model paling baik dengan nilai *accuracy* sebesar 80%. Sedangkan hasil pengujian dengan menggunakan *Naive Bayes* saja, hanya dapat memberikan performa model dengan nilai *accuracy* sebesar 66%. Hal ini dikarenakan data set yang digunakan pada skenario pengujian 4, sudah seimbang pada setiap kelasnya dan memiliki kriteria yang relevan. Sehingga nilai-nilai probabilitas dari kriteria tersebut dapat dioptimalkan untuk memprediksi suatu kondisi berdasarkan proses klasifikasi.

5.2 Saran

Adapun hal-hal yang menjadi saran sebagai bahan pertimbangan untuk pengembangan penelitian selanjutnya agar tahapan dalam melakukan penelitian menjadi lebih efektif dan model yang dihasilkan menjadi lebih baik adalah sebagai berikut:

1. Mengkombinasikan data terkait klasifikasi bidang keahlian dengan menggunakan indeks Prestasi Kumulatif Semester 5 dan 6, minat mahasiswa, dan kemampuan mahasiswa (*hard skill*).
2. Menggunakan metode lainnya pada tahapan *preprocessing* dan *feature selection* agar dapat dilakukan perbandingan serta dapat dilihat perbedaan hasil performa algoritma klasifikasi *Naive Bayes* dan memperoleh nilai akurasi dari masing-masing metode yang digunakan.
3. Melakukan klasifikasi dengan algoritma pengklasifikasian lainnya untuk mendapatkan performa dari masing-masing algoritma yang digunakan.

DAFTAR PUSTAKA

- Advernesia. (2017, Mei). *Apa itu Machine Learning dan Cara Kerjanya*. Diambil kembali dari Advernesia Matematika Komputer Internet: <https://www.advernesia.com/blog>
- Arddy H. Hailitik, B. S. (2017). KLASIFIKASI JURUSAN MENGGUNAKAN METODE NAÏVE BAYES PADA SEKOLAH MENENGAH ATAS NEGERI (SMAN) 1 FATULEU TENGAH. *Jicon*, 21-27.
- ARWAN, V. A. (2018, Juni 8). *Synthetic Minority Over-sampling Technique (SMOTE) Algorithm For Handling Imbalanced Data*. Diambil kembali dari BINUS UNIVERSITY: <https://mti.binus.ac.id/>
- Bisri, M. H. (2015). Implementasi Algoritma Naïve Bayes untuk Memprediksi Penjurusan Siswa di SMA Kesatrian 1 Semarang. *Jurusan Teknik Informatika, FASILKOM UDINUS Jln. Nakula 1 No 5-11 Semarang 50131 INDONESIA*.
- Christobel, A. & Sivaprakasam, D., 2011. An Empirical Comparison of Data Mining Classification Methods. , 3(2), pp.24–28.
- Darmawan, W. (2017). ALGORITMA KLASIFIKASI DATA MINING UNTUK PREDIKSI STATUS MAHASISWA MENGGUNAKAN ALGORITMA DECISION TREE (C4.5). *IC-Tech*, 17.
- Deni Irvantoro, 2. S. (2019). Feature Selection Menggunakan Chi-Square Dan N-Gram Dengan Algoritma Naive Bayes Classifier Untuk Analisis Sentimen Review Produk Elektronik. 3.
- dsn, c. (2018, April 13). *Naïve Bayes Classifier (NBC)*. Diambil kembali dari Cahya's Blog: https://cahyadsn.phpindonesia.id/extra/naive_bayes.php
- Fajar, P. S. (2020). Analisis sentimen komentar mahasiswa terhadap kinerja dosen menggunakan metode multinomial naïve bayes dan bernoulli naïve bayes . 20.
- Fatayat, R. A. (2021). ANALISA PENENTUAN DOSEN PEMBIMBING TUGAS AKHIR MAHASISWA MENGGUNAKAN NAIVE BAYES CLASSIFIER. *SIMTIKA* , 3.
- Faz Faidhani, T. A. (2021). Sistem Pendukung Keputusan Penentu Bidang Keahlian Mahasiswa Program Studi Informatika Universitas Tanjungpura dengan Metode ELECTRE. *JUSTIN*, 42.
- Fuad, N. (2019). ALGORITMA FUZZY NAIVE BAYES UNTUK MENGLASIFIKASIKAN BIDANG KEAHLIAN MAHASISWA TEKNIK INFORMATIKA UNIVERSITAS ISLAM LAMONGAN. *JOUTICA*, 302.

- Hastuti, K., 2012. Analisis Komparasi Algoritma Klasifikasi Data Mining untuk Prediksi Mahasiswa Non-aktif. , 2012(Semantik 2012), pp.241–249.
- Husin Muhamad, C. A. (2017). OPTIMASI NAIVE BAYES CLASSIFIER DENGAN MENGGUNAKAN PARTICLE. *JTIK*, 180.
- Igo Cahya Negara, A. P. (2018). PENGGUNAAN UJI CHI-SQUARE UNTUK MENGETAHUI PENGARUH TINGKAT PENDIDIKAN DAN UMUR TERHADAP PENGETAHUAN. *Prosiding Seminar Nasional Matematika dan Terapannya*, 3.
- Komputer, K. (2018, July 23). *Pengujian Dengan Confusion Matrix*. Diambil kembali dari Kuliah Komputer: <http://www.kuliahkomputer.com/2018/07/pengujian-dengan-confusion-matrix.html>
- Kotsiantis, S.B., Pierrakeas, C.J. & Pintelas, P.E., 2003. Preventing Student Dropout in Distance Learning Using Machine Learning Techniques. , pp.267–274.
- Laila Qadrini, A. S. (2021). DECISION TREE DAN ADABOOST PADA KLASIFIKASI PENERIMA PROGRAM BANTUAN SOSIAL. *JIP*, 1962.
- Medina, M. I. (2021, Agustus 26). *Penting untuk Data Scientist, Ketahui Apa Saja Fungsi dan Fitur Jupyter*. Diambil kembali dari glints: <https://glints.com>
- Muslich, M. (2008, Maret 23). *Apa itu KTI?* Diambil kembali dari <http://muslich-m-blogspot.com/2008/03/apa-itu-kti-html>
- Pendahuluan Python*. (t.thn.). Diambil kembali dari Belajarpython: <https://belajarpython.com/tutorial/apa-itu-python>
- Punlumjeak, Wattana, Rachburee, Nachirat, 2015, A Comparative Study of Feature Selection Techniques for Classify Student Performance, Chiang Mai, Thailand.
- Ragab, A.H.M. et al., 2014. A Comparative Analysis of Classification Algorithms for Students College Enrollment Approval Using Data Mining. Proceedings of the 2014 Workshop on Interaction Design in Educational Environments - IDEE '14, pp.106–113.
- Siringoringo, R., 2018, Klasifikasi Data Tidak Seimbang menggunakan Algoritma SMOTE dan K- Nearest Neighbor, Jurnal ISD.
- SOFANUDIN, A. (2017). IMPLEMENTASI ALGORITMA NAIVE BAYES UNTUK PENJURUSAN DI MA AL MAHRUSIYAH LIRBOYO. *Artikel Skripsi Universitas Nusantara PGRI Kediri*, 3.
- Swastina, L. (2013). Penerapan Algoritma C4.5 Untuk Penentuan. *GEMA AKTUALITA*, 93-98.

- Tentang Prodi.* (2016). Diambil kembali dari Teknik Informatika Fakultas Teknik Universitas Tanjungpura: <http://informatika.untan.ac.id/?single&id=135>
- Wede. (2020, November 17). *3 Library yang Esensial dalam Belajar Machine Learning dengan Python.* Diambil kembali dari DQLab: <https://www.dqlab.id>
- Widianto, M. H. (2019, Desember 23). *Algoritma Naive Bayes.* Diambil kembali dari BINUS UNIVERSITY: <https://binus.ac.id>
- Wihardi, Y. (2013, April 2). Diambil kembali dari K-Folds Cross Validation: <http://blog.yayaw.web.id/riset/k-folds-cross-validation>
- Witten, I.H., Frank, E. & Hall, M. a., 2011. Data Mining: Practical Machine Learning Tools and Techniques Third Edit., USA: Morgan Kaufmann Publishers.

LAMPIRAN A

KUESIONER REQUIREMENT KEMAMPUAN INFORMATIKA

Requirement Kemampuan Informatika

Kirim

Pertanyaan Jawaban 20 Setelan Poin total: 0

Bagian 1 dari 2

Requirement Kemampuan

Kepada Yth :
Bapak/Ibu responden

Saya Fajar Istighfar Mahasiswa Jurusan Informatika dengan NIM D1041161002. Saat ini saya sedang melaksanakan Tugas Akhir dengan judul Klasifikasi Bidang Keahlian Pada Mahasiswa Informatika Fakultas Teknik Universitas Tanjungpura Menggunakan Algoritma Naive Bayes. Dalam rangka pengumpulan data, saya memohon partisipasi dari Bapak/Ibu selaku ketua dan anggota kelompok keahlian di Informatika Fakultas Teknik Universitas Tanjungpura, agar dapat mengisi kuesioner ini. Kuesioner berisi tentang kemampuan apa saja yang harus dikuasai oleh mahasiswa terkait kelompok keahlian. Adapun Kemampuan yang harus dikuasai oleh Mahasiswa pada Kelompok Keahlian, saya bagi menjadi :

1. Penguasaan Bahasa Pemrograman
2. Penguasaan Tools/Aplikasi
3. Pemahaman dan Penerapan Konsep

Data yang terkumpul dari kuesioner ini akan saya sampaikan ke Bapak/Ibu, dan akan saya gunakan untuk penelitian tugas akhir saya. Atas perhatian dan waktunya Bapak/Ibu saya ucapkan terimakasih.

Hormat Saya,
Fajar Istighfar

Formulir ini mengumpulkan email untuk pengguna Universitas Tanjungpura secara otomatis. [Ubah setelan](#)

Kelompok Keahlian

☐ Kelompok Keahlian Sistem Informasi dan Data Spasial

☐ Kelompok Keahlian Rekayasa Perangkat Lunak dan Komputasi Bergerak

☐ Kelompok Keahlian Jaringan dan Keamanan

☐ Kelompok Keahlian Komputasi dan Kecerdasan Buatan

Setelah bagian 1 Lanjutkan ke bagian berikut

Bagian 2 dari 2

Kemampuan yang harus dikuasai oleh Mahasiswa pada Kelompok Keahlian

Setiap Kemampuan mempunyai beberapa pertanyaan dan wajib diisi di setiap kemampuan nya.
Petunjuk Pengisian
berikan centang pada kolom
Kolom 1: Tidak Diperlukan
Kolom 2: Cukup Diperlukan
Kolom 3: Diperlukan
Kolom 4: Sangat Diperlukan
Adapun Kemampuan yang harus dikuasai oleh Mahasiswa pada Kelompok Keahlian terbagi menjadi :

1. Penguasaan Bahasa Pemrograman
2. Penguasaan Tools/Aplikasi
3. Pemahaman dan Penerapan Konsep

1. Penguasaan Bahasa Pemrograman

	Tidak Diperlukan	Cukup Diperlukan	Diperlukan	Sangat Diperlukan
a. Bahasa pemrogr...	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
b. Bahasa pemrogr...	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
c. Bahasa pemrogr...	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
d. Bahasa pemrogr...	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
e. Bahasa pemrogr...	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
f. Bahasa pemrogr...	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
g. Bahasa pemrogr...	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
h. Bahasa pemrogr...	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
i. Bahasa pemrogra...	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
j. Bahasa pemrogra...	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
k. Bahasa pemrogr...	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
l. Bahasa pemrogra...	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
m. Bahasa pemrog...	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
n. Bahasa pemrogr...	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
o. Bahasa pemrogr...	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

A-1

LAMPIRAN B

KUESIONER KEMAMPUAN MAHASISWA

Requirement Kemampuan Mahasiswa Informatika

PertanyaanJawaban139Setelan

Requirement Kemampuan Mahasiswa Informatika

Kepada Yth :
Mahasiswa responden

Saya Fajar Istighfar Mahasiswa Jurusan Informatika dengan NIM D1041161002. Saat ini saya sedang melaksanakan Tugas Akhir dengan judul Klasifikasi Bidang Keahlian Pada Mahasiswa Informatika Fakultas Teknik Universitas Tanjungpura Menggunakan Algoritma Naive Bayes. Dalam rangka pengumpulan data, saya memohon partisipasi dari saudara/i selaku mahasiswa di Informatika Fakultas Teknik Universitas Tanjungpura, agar dapat mengisi kuesioner ini. Kuesioner berisi tentang kemampuan apa saja yang harus dikuasai oleh mahasiswa terkait kelompok keahlian. Adapun Kemampuan yang harus dikuasai oleh Mahasiswa pada Kelompok Keahlian, saya bagi menjadi :
1. Penguasaan Bahasa Pemrograman
2. Penguasaan Tools/Aplikasi
3. Pemahaman dan Penerapan Konsep

Data yang terkumpul dari kuesioner ini akan saya gunakan untuk penelitian tugas akhir saya. Atas perhatian dan waktunya saya ucapkan terimakasih.

Hormat Saya,
Fajar Istighfar

Formulir ini mengumpulkan email untuk pengguna Universitas Tanjungpura secara otomatis. [Ubah setelan](#)

IDENTITAS RESPONDEN

Deskripsi (opsional)

Nama *

Teks jawaban singkat

Nim *

Teks jawaban singkat

Tingkat kemampuan yang harus dikuasai oleh Mahasiswa pada Kelompok Keahlian

Keterangan dan petunjuk pengisian.
Pada bagian ini anda diminta untuk mengisi tingkat kemampuan yang anda kuasai, setiap kemampuan mempunyai beberapa pertanyaan dan wajib diisi dengan memberikan centang pada kolom di setiap kemampuan nya.
Kolom 1 : Tidak Bisa
Kolom 2 : Bisa

1. Penguasaan Bahasa Pemrograman *

	Tidak Bisa	Bisa
a. Bahasa pemrograman XML	<input type="radio"/>	<input type="radio"/>
b. Bahasa perograman Java	<input type="radio"/>	<input type="radio"/>
c. Bahasa pemrograman Javascript	<input type="radio"/>	<input type="radio"/>
d. Bahasa pemrograman PHP dan...	<input type="radio"/>	<input type="radio"/>
e. Bahasa pemrograman HTML	<input type="radio"/>	<input type="radio"/>
f. Bahasa pemrograman SQL	<input type="radio"/>	<input type="radio"/>
g. Bahasa pemrograman Prolog	<input type="radio"/>	<input type="radio"/>
h. Bahasa pemrograman RUBY	<input type="radio"/>	<input type="radio"/>
i. Bahasa pemrograman Python	<input type="radio"/>	<input type="radio"/>
j. Bahasa pemrograman Curl	<input type="radio"/>	<input type="radio"/>
k. Bahasa pemrograman COBOL	<input type="radio"/>	<input type="radio"/>
l. Bahasa pemrograman PERL	<input type="radio"/>	<input type="radio"/>
m. Bahasa pemrograman Swift	<input type="radio"/>	<input type="radio"/>
n. Bahasa pemrograman Visual B...	<input type="radio"/>	<input type="radio"/>
o. Bahasa pemrograman C atau C...	<input type="radio"/>	<input type="radio"/>

B-1

LAMPIRAN C

HASIL KUESIONER REQUIREMENT KEMAMPUAN INFORMATIKA

Time stamp	Email Address	Kelompok Keahlian	1. Penguasaan Bahasa Pemrograman [a. Bahasa pemrograman XML]	1. Penguasaan Bahasa Pemrograman [b. Bahasa pemrograman Java]	1. Penguasaan Bahasa Pemrograman [c. Bahasa pemrograman Javascript]	...	3. Pemahaman dan Penerapan Konsep [a. Pemahaman konsep XHTML]	3. Pemahaman dan Penerapan Konsep [b. Pemahaman dasar dan konsep Desain Grafis]	3. Pemahaman dan Penerapan Konsep [c. Pemahaman konsep Monitor dan administrasi Network Security]
4/21/2021 8:33:03	helensastypratiwi@informatika.cs.untan.ac.id	Kelompok Keahlian Komputasi dan Kecerdasan Buatan	Diperlukan	Diperlukan	Diperlukan	...	Diperlukan	Sangat Diperlukan	Cukup Diperlukan
4/21/2021 8:34:46	safriadi@informatics.untan.ac.id	Kelompok Keahlian Rekayasa Perangkat Lunak dan Komputasi Bergerak	Diperlukan	Sangat Diperlukan	Sangat Diperlukan	...	Sangat Diperlukan	Sangat Diperlukan	Tidak Diperlukan
4/21/2021 16:32:20	irwansyah.azhar@untan.ac.id	Kelompok Keahlian Jaringan dan Keamanan	Cukup Diperlukan	Cukup Diperlukan	Cukup Diperlukan	...	Cukup Diperlukan	Tidak Diperlukan	Diperlukan

Time stamp	Email Address	Kelompok Keahlian	1. Penguasaan Bahasa Pemrograman [a. Bahasa pemrograman XML]	1. Penguasaan Bahasa Pemrograman [b. Bahasa pemrograman Java]	1. Penguasaan Bahasa Pemrograman [c. Bahasa pemrograman Javascript]	...	3. Pemahaman dan Penerapan Konsep [a. Pemahaman konsep XHTML]	3. Pemahaman dan Penerapan Konsep [b. Pemahaman dasar dan konsep Desain Grafis]	3. Pemahaman dan Penerapan Konsep [c. Pemahaman konsep Monitor dan administrasi Network Security]
6/11/2021 8:50:54	anggiperwitasari@informatika.untan.ac.id	Kelompok Keahlian Rekayasa Perangkat Lunak dan Komputasi Bergerak	Sangat Diperlukan	Sangat Diperlukan	Sangat Diperlukan	...	Sangat Diperlukan	Sangat Diperlukan	Diperlukan
6/11/2021 8:59:26	sholvariza@untan.ac.id	Kelompok Keahlian Sistem Informasi dan Data Spasial	Cukup Diperlukan	Cukup Diperlukan	Cukup Diperlukan	...	Sangat Diperlukan	Sangat Diperlukan	Diperlukan
6/11/2021 9:34:51	haried@informatika.untan.ac.id	Kelompok Keahlian Jaringan dan Keamanan	Cukup Diperlukan	Cukup Diperlukan	Diperlukan	...	Cukup Diperlukan	Cukup Diperlukan	Diperlukan
6/11/2021 10:22:06	rinaseptiriana@informatika.untan.ac.id	Kelompok Keahlian Komputasi dan Kecerdasan Buatan	Tidak Diperlukan	Cukup Diperlukan	Tidak Diperlukan	...	Cukup Diperlukan	Tidak Diperlukan	Tidak Diperlukan

Time stamp	Email Address	Kelompok Keahlian	1. Penguasaan Bahasa Pemrograman [a. Bahasa pemrograman XML]	1. Penguasaan Bahasa Pemrograman [b. Bahasa pemrograman Java]	1. Penguasaan Bahasa Pemrograman [c. Bahasa pemrograman Javascript]	...	3. Pemahaman dan Penerapan Konsep [a. Pemahaman konsep XHTML]	3. Pemahaman dan Penerapan Konsep [b. Pemahaman dasar dan konsep Desain Grafis]	3. Pemahaman dan Penerapan Konsep [c. Pemahaman konsep Monitor dan administrasi Network Security]
6/14/2021 13:09:24	helfinas@informatika.untan.ac.id	Kelompok Keahlian Jaringan dan Keamanan	Sangat Diperlukan	Sangat Diperlukan	Sangat Diperlukan	...	Sangat Diperlukan	Sangat Diperlukan	Sangat Diperlukan
6/14/2021 17:41:52	morteza.muthahhari@teknik.untan.ac.id	Kelompok Keahlian Sistem Informasi dan Data Spasial	Cukup Diperlukan	Diperlukan	Diperlukan	...	Cukup Diperlukan	Cukup Diperlukan	Tidak Diperlukan
6/16/2021 11:32:42	rudydn@informatika.untan.ac.id	Kelompok Keahlian Rekayasa Perangkat Lunak dan Komputasi Bergerak	Cukup Diperlukan	Cukup Diperlukan	Cukup Diperlukan	...	Cukup Diperlukan	Cukup Diperlukan	Tidak Diperlukan
6/21/2021 8:56:23	hs@untan.ac.id	Kelompok Keahlian Komputasi dan Kecerdasan Buatan	Diperlukan	Diperlukan	Diperlukan	...	Diperlukan	Diperlukan	Cukup Diperlukan

Time stamp	Email Address	Kelompok Keahlian	1. Penguasaan Bahasa Pemrograman [a. Bahasa pemrograman XML]	1. Penguasaan Bahasa Pemrograman [b. Bahasa pemrograman Java]	1. Penguasaan Bahasa Pemrograman [c. Bahasa pemrograman Javascript]	...	3. Pemahaman dan Penerapan Konsep [a. Pemahaman konsep XHTML]	3. Pemahaman dan Penerapan Konsep [b. Pemahaman dasar dan konsep Desain Grafis]	3. Pemahaman dan Penerapan Konsep [c. Pemahaman konsep Monitor dan administrasi Network Security]
6/21/2021 10:09:11	anggidianti@informatika.untan.ac.id	Kelompok Keahlian Sistem Informasi dan Data Spasial	Diperlukan	Diperlukan	Diperlukan	...	Diperlukan	Diperlukan	Tidak Diperlukan
6/24/2021 23:18:53	tursina@informatika.untan.ac.id	Kelompok Keahlian Komputasi dan Kecerdasan Buatan	Cukup Diperlukan	Diperlukan	Diperlukan	...	Cukup Diperlukan	Diperlukan	Cukup Diperlukan
6/25/2021 21:09:19	hengkyanra@informatika.untan.ac.id	Kelompok Keahlian Rekayasa Perangkat Lunak dan Komputasi Bergerak	Sangat Diperlukan	Sangat Diperlukan	Sangat Diperlukan	...	Sangat Diperlukan	Sangat Diperlukan	Sangat Diperlukan
6/28/2021 18:51:08	arifbpn@untan.ac.id	Kelompok Keahlian Komputasi dan Kecerdasan Buatan	Cukup Diperlukan	Sangat Diperlukan	Cukup Diperlukan	...	Tidak Diperlukan	Tidak Diperlukan	Tidak Diperlukan

Time stamp	Email Address	Kelompok Keahlian	1. Penguasaan Bahasa Pemrograman [a. Bahasa pemrograman XML]	1. Penguasaan Bahasa Pemrograman [b. Bahasa perograman Java]	1. Penguasaan Bahasa Pemrograman [c. Bahasa pemrograman Javascript]	...	3. Pemahaman dan Penerapan Konsep [a. Pemahaman konsep XHTML]	3. Pemahaman dan Penerapan Konsep [b. Pemahaman dasar dan konsep Desain Grafis]	3. Pemahaman dan Penerapan Konsep [c. Pemahaman konsep Monitor dan administrasi Network Security]
7/5/2021 6:14:20	media.ft@untan.ac.id	Kelompok Keahlian Komputasi dan Kecerdasan Buatan	Diperlukan	Diperlukan	Diperlukan	...	Cukup Diperlukan	Cukup Diperlukan	Cukup Diperlukan
7/5/2021 7:38:12	evaripanti@untan.ac.id	Kelompok Keahlian Sistem Informasi dan Data Spasial	Diperlukan	Diperlukan	Diperlukan	...	Diperlukan	Diperlukan	Diperlukan

LAMPIRAN D

HASIL KUESIONER KEMAMPUAN MAHASISWA

Time stamp	Email Address	Nama	Nim	1. Penguasaan Bahasa Pemrograman [a. Bahasa pemrograman XML]	1. Penguasaan Bahasa Pemrograman [b. Bahasa pemrograman Java]	1. Penguasaan Bahasa Pemrograman [c. Bahasa pemrograman Javascript]	...	3. Pemahaman dan Penerapan Konsep [a. Pemahaman konsep XHTML]	3. Pemahaman dan Penerapan Konsep [b. Pemahaman dasar dan konsep Desain Grafis]	3. Pemahaman dan Penerapan Konsep [c. Pemahaman konsep Monitor dan administrasi Network Security]
8/10/2021 16:20:13	windaoktapias@student.untan.ac.id	Windari Oktapia Simanjuntak	D1041161032	Tidak Bisa	Tidak Bisa	Tidak Bisa	...	Tidak Bisa	Bisa	Tidak Bisa
8/10/2021 16:21:17	bravobb@student.untan.ac.id	Ansfridus Bravo	D1041161024	Tidak Bisa	Tidak Bisa	Tidak Bisa	...	Tidak Bisa	Bisa	Tidak Bisa
8/10/2021 16:23:11	rama.ulgasesa@student.untan.ac.id	Rama Ulgasesa	D1041161026	Tidak Bisa	Tidak Bisa	Tidak Bisa	...	Tidak Bisa	Bisa	Tidak Bisa
8/10/2021 16:26:20	farhan.hersinanda@student.untan.ac.id	Muhammad Farhan Hersinanda	D1041181026	Tidak Bisa	Tidak Bisa	Tidak Bisa	...	Tidak Bisa	Bisa	Tidak Bisa
8/10/2021 16:31:58	adityawijaya@student.untan.ac.id	Aditya Wijaya	D1041181038	Tidak Bisa	Bisa	Bisa	...	Tidak Bisa	Bisa	Bisa
8/10/2021 16:32:53	enrikogreen@student.untan.ac.id	Enriko Yudhistira Ramadhan	D1041161004	Tidak Bisa	Bisa	Bisa	...	Tidak Bisa	Bisa	Tidak Bisa
8/10/2021 16:35:01	kusmayuda@student.untan.ac.id	Kusmayuda	D1041181014	Tidak Bisa	Tidak Bisa	Bisa	...	Bisa	Bisa	Tidak Bisa
8/10/2021 16:35:03	zanuardydwinata@student.untan.ac.id	Zanuardy Supria Dwinata	D1041171037	Bisa	Tidak Bisa	Tidak Bisa	...	Bisa	Tidak Bisa	Tidak Bisa

Time stamp	Email Address	Nama	Nim	1. Penguasaan Bahasa Pemrograman [a. Bahasa pemrograman XML]	1. Penguasaan Bahasa Pemrograman [b. Bahasa perograman Java]	1. Penguasaan Bahasa Pemrograman [c. Bahasa pemrograman Javascript]	...	3. Pemahaman dan Penerapan Konsep [a. Pemahaman konsep XHTML]	3. Pemahaman dan Penerapan Konsep [b. Pemahaman dasar dan konsep Desain Grafis]	3. Pemahaman dan Penerapan Konsep [c. Pemahaman konsep Monitor dan administrasi Network Security]
8/10/2021 16:35:08	ifan.fathoni@student.untan.ac.id	Ifan Fathoni Anshari	D1041161040	Tidak Bisa	Bisa	Bisa	...	Tidak Bisa	Bisa	Bisa
8/10/2021 16:36:11	ihsanz10@student.untan.ac.id	Ihsan Maulana	D1041161016	Bisa	Tidak Bisa	Bisa	...	Bisa	Tidak Bisa	Tidak Bisa
...
10/15/2021 21:27:35	rianmkc17@student.untan.ac.id	Rian Fernando	D1042171041	Tidak Bisa	Tidak Bisa	Tidak Bisa	...	Bisa	Bisa	Bisa

LAMPIRAN E PELABELAN

Nama	Nim	1. Penguasaan Bahasa Pemrograman [a. Bahasa pemrograman XML]	1. Penguasaan Bahasa Pemrograman [b. Bahasa pemrograman Java]	1. Penguasaan Bahasa Pemrograman [c. Bahasa pemrograman Javascript]	...	3. Pemahaman dan Penerapan Konsep [a. Pemahaman konsep XHTML]	3. Pemahaman dan Penerapan Konsep [b. Pemahaman dasar dan konsep Desain Grafis]	3. Pemahaman dan Penerapan Konsep [c. Pemahaman konsep Monitor dan administrasi Network Security]	Label 1
Windari Oktapia Simanjuntak	D1041161032	Tidak Bisa	Tidak Bisa	Tidak Bisa	...	Tidak Bisa	Bisa	Tidak Bisa	Kelompok Keahlian Komputasi dan Kecerdasan Buatan
Ansfridus Bravo	D1041161024	Tidak Bisa	Tidak Bisa	Tidak Bisa	...	Tidak Bisa	Bisa	Tidak Bisa	Kelompok Keahlian Rekayasa Perangkat Lunak dan Komputasi Bergerak
Rama Ulgasesa	D1041161026	Tidak Bisa	Tidak Bisa	Tidak Bisa	...	Tidak Bisa	Bisa	Tidak Bisa	Kelompok Keahlian Komputasi dan Kecerdasan Buatan
Muhammad Farhan Hersinanda	D1041181026	Tidak Bisa	Tidak Bisa	Tidak Bisa	...	Tidak Bisa	Bisa	Tidak Bisa	Kelompok Keahlian Sistem Informasi dan Data Spasial
Aditya Wijaya	D1041181038	Tidak Bisa	Bisa	Bisa	...	Tidak Bisa	Bisa	Bisa	Kelompok Keahlian Jaringan dan Keamanan
Enriko Yudhistira Ramadhan	D1041161004	Tidak Bisa	Bisa	Bisa	...	Tidak Bisa	Bisa	Tidak Bisa	Kelompok Keahlian Jaringan dan Keamanan

Nama	Nim	1. Penguasaan Bahasa Pemrograman [a. Bahasa pemrograman XML]	1. Penguasaan Bahasa Pemrograman [b. Bahasa pemrograman Java]	1. Penguasaan Bahasa Pemrograman [c. Bahasa pemrograman Javascript]	...	3. Pemahaman dan Penerapan Konsep [a. Pemahaman konsep XHTML]	3. Pemahaman dan Penerapan Konsep [b. Pemahaman dasar dan konsep Desain Grafis]	3. Pemahaman dan Penerapan Konsep [c. Pemahaman konsep Monitor dan administrasi Network Security]	Label 1
Kusmayuda	D1041181014	Tidak Bisa	Tidak Bisa	Bisa	...	Bisa	Bisa	Tidak Bisa	Kelompok Keahlian Sistem Informasi dan Data Spasial
Zanuary Supria Dwinata	D1041171037	Bisa	Tidak Bisa	Tidak Bisa	...	Bisa	Tidak Bisa	Tidak Bisa	Kelompok Keahlian Rekayasa Perangkat Lunak dan Komputasi Bergerak
Ifan Fathoni Anshari	D1041161040	Tidak Bisa	Bisa	Bisa	...	Tidak Bisa	Bisa	Bisa	Kelompok Keahlian Sistem Informasi dan Data Spasial
Ihsan Maulana	D1041161016	Bisa	Tidak Bisa	Bisa	...	Bisa	Tidak Bisa	Tidak Bisa	Kelompok Keahlian Sistem Informasi dan Data Spasial
...
Rian Fernando	D1042171041	Tidak Bisa	Tidak Bisa	Tidak Bisa	...	Bisa	Bisa	Bisa	Kelompok Keahlian Sistem Informasi dan Data Spasial

LAMPIRAN F HASIL *CONFUSION MATRIX*

Skenario Pertama

Confusion Matrix:

```
[[47  1  8  1]
 [ 8  6  3  1]
 [13  2 29  2]
 [ 1  1  3  2]]
```

Classification Report:

	precision	recall	f1-score	support
0	0.68	0.82	0.75	57
1	0.60	0.33	0.43	18
2	0.67	0.63	0.65	46
3	0.33	0.29	0.31	7
accuracy			0.66	128
macro avg	0.57	0.52	0.53	128
weighted avg	0.65	0.66	0.64	128

Skenario Kedua

Confusion Matrix:

```
[[48  2  5  2]
 [ 5  9  2  2]
 [12  1 30  3]
 [ 0  2  1  4]]
```

Classification Report:

	precision	recall	f1-score	support
0	0.74	0.84	0.79	57
1	0.64	0.50	0.56	18
2	0.79	0.65	0.71	46
3	0.36	0.57	0.44	7
accuracy			0.71	128
macro avg	0.63	0.64	0.63	128
weighted avg	0.72	0.71	0.71	128

Skenario Ketiga

Confusion Matrix:

```
[[39  8  8  2]
 [ 5 47  3  2]
 [10  6 40  1]
 [ 0  0  3 54]]
```

Classification Report:

	precision	recall	f1-score	support
0	0.72	0.68	0.70	57
1	0.77	0.82	0.80	57
2	0.74	0.70	0.72	57
3	0.92	0.95	0.93	57
accuracy			0.79	228
macro avg	0.79	0.79	0.79	228
weighted avg	0.79	0.79	0.79	228

Skenario Keempat

Confusion Matrix:

```
[[40  7  9  1]
 [ 4 48  3  2]
 [ 9  6 41  1]
 [ 0  0  3 54]]
```

Classification Report:

	precision	recall	f1-score	support
0	0.75	0.70	0.73	57
1	0.79	0.84	0.81	57
2	0.73	0.72	0.73	57
3	0.93	0.95	0.94	57
accuracy			0.80	228
macro avg	0.80	0.80	0.80	228
weighted avg	0.80	0.80	0.80	228