



**KLASIFIKASI PENERIMA KIP SMPN 5 UNGARAN
MENGUNAKAN METODE NAÏVE BAYES
BERBASIS WEB**

Oleh :

**NAMA : DINIA KURNIA PRATIWI (A11.2016.09772)
MUHAMMAD HIZBULLAH(A11.2016.10006)**

Program Studi : Teknik Informatika-S1

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS DIAN NUSWANTORO
SEMARANG
TAHUN 2019**

DAFTAR ISI

BAB I

PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah	2
1.4 Tujuan Penelitian	2
1.5 Manfaat Penelitian	3
1.5.1 Bagi SMPN 5 Ungaran	3
1.5.2 Bagi Universitas Dian Nuswantoro	3
1.5.3 Bagi Pembaca	3

BAB II

TINJAUAN PUSTAKA	4
2.1 Tinjauan Studi	4
2.2 Tinjauan Pustaka	11
2.2.1 Klasifikasi	11
2.2.2 Data Mining	12
Gambar 2.1 Tahap-tahap Data Mining	12
1. Pembersihan data (data cleaning)	12
2. Integrasi data (data integration)	13
3. Seleksi Data (Data Selection)	13
4. Transformasi data (Data Transformation)	13
5. Proses mining,	13
6. Evaluasi pola (pattern evaluation),	14

7. Presentasi pengetahuan (knowledge presentation),.....	14
Gambar 2.2 Crisp-Data Mining	15
1. Business Understanding	15
2. Data Understanding.....	16
3. Data Preparation.....	16
4. Modeling	16
5. Evaluation	17
6. Deployment	17
2.2.3 Klasifikasi.....	17
2.2.4 Naïve Bayes Classifier	18
2.2.5 UML.....	19
Gambar 2.3 Use Case Diagram Sistem	19
2.2.6 Model Pengembangan Sistem	20
Waterfall.....	20
Gambar 2.4 Metode Waterfall.....	20
2.3 Kerangka Pemikiran	22
Gambar 2.5 Fishbone Diagram	22
BAB III	23
MEDOTE PENELITIAN	23
3.1 Metode Pengumpulan Data	23
3.3 Model Yang Diusulkan	25
3.3.1 Modelling	25
3.3.2 Naïve Bayes Classifier	26
3.4 Pengujian Metode Naive Bayes.....	27
BAB IV	28
JADWAL PENELITIAN DATA MINING	28
4.1 Jadwal Penyusunan Penelitian Data Mining	28

DAFTAR GAMBAR

Gambar 2.1 Tahap-tahap Data Mining.....	12
Gambar 2.2 Crisp-Data Mining	15
Gambar 2.3 Use Case Diagram Sistem	19
Gambar 2.4 Metode Waterfall.....	20
Gambar 2.5 Fishbone Diagram	22
Gambar 3.1 Flowchart Untuk Menemukan Performance Algoritma	24
Gambar 3.2 Flowchart Sistem Dengan Algoritma Naive Bayes	25

DAFTAR TABEL

Tabel 2.1 State Of The Art	6
Tabel 3.1 Detail Atribut Data Penelitian	23
Tabel 3.2 Atribut Data Yang Digunakan	23
Tabel 4.1 Jadwal Penyusunan Penelitian Data Mining	28

BAB I

PENDAHULUAN

1.1 Latar Belakang

KIP (Kartu Indonesia Pintar) merupakan kartu yang ditujukan bagi keluarga miskin dan rentan miskin yang ingin menyekolahkan anak yang berusia 7-18 tahun secara gratis. Mereka yang mendapat KIP ini akan diberikan dana tunai dari pemerintah secara reguler yang tersimpan dalam fungsi kartu KIP untuk bersekolah secara gratis tanpa biaya. Program KIP sendiri akan ditujukan pada 15,5 juta keluarga kurang mampu di seluruh Indonesia yang memiliki anak usia sekolah 7 hingga 18 tahun baik yang telah terdaftar maupun yang belum terdaftar di sekolah maupun madrasah. Dengan program KIP ini diharapkan angka putus sekolah bisa turun dengan drastis.[1]

Penerima program KIP ini sendiri diprioritaskan pada :

- Penerima BSM dari Pemegang KPS yang telah ditetapkan dalam SP2D 2014
- Anak usia sekolah (6-21 tahun) dari keluarga pemegang KPS/KKS yang belum ditetapkan sebagai Penerima manfaat BSM
- Anak usia sekolah (6-21 tahun) dari keluarga peserta PKH
- Anak usia sekolah (6-21 tahun) yang tinggal di Panti Asuhan/Sosial
- Siswa/santri (6-21 tahun) dari Pondok Pesantren yang memiliki KPS/KKS (khusus untuk BSM Mandrasah)
- Anak usia sekolah (6-21 tahun) yang terancam putus sekolah karena kesulitan ekonomi dan/atau korban musibah berkepanjangan/bencana alam melalui jalur FUS/FUM;
- Anak usia sekolah yang belum atau tidak lagi bersekolah yang datanya telah dapat direkapitulasi pada Semester 2 (TA) 2014/2015.

Dari uraian kasus diatas, untuk mengelompokkan sebuah layak tidaknya untuk mendapatkan KIP dapat menggunakan klasifikasi Naïve Bayes data mining. Proses klasifikasi merupakan sebuah proses penemuan model atau fungsi yang dapat menjelaskan atau membedakan konsep data, yang bertujuan untuk dapat memperkirakan konsep tersebut dari suatu objek yang labelnya tidak diketahui [2]. Data mining adalah proses untuk memperoleh keterkaitan dalam sebuah data yang belum diketahui, kemudian disajikan dengan suatu cara yang dapat dipahami oleh pengguna sehingga keterkaitan tersebut dapat dijadikan sebagai dasar pengambilan

keputusan [3]. Metode Naive Bayes dipilih karena hasil lebih akurat dibanding metode C4.5 dan teknik lebih sederhana, hal ini dibuktikan oleh Doreswamy, Hemanth. K. S dalam jurnalnya tahun 2012 “Performance Evaluation of Predictive Classifiers For Knowledge Discovery

From Engineering Materials Data Sets”.

Dengan adanya klasifikasi penerima KIP menggunakan metode naive bayes diharapkan dapat membantu pihak sekolah dalam mengelompokkan Kelayakan penerima KIP .

1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, dapat dirumuskan permasalahan yang akan diselesaikan adalah “Bagaimana kinerja naïve bayes dalam pengklasifikasian penerima KIP di SMPN 5 Ungaran dan menerapkannya ke dalam aplikasi data mining? “.

1.3 Batasan Masalah

Sesuai dengan judul dan tujuan yang diangkat, maka penulis akan membuat ruang lingkup dan batasan masalah yaitu :

1. Pengumpulan data siswa SMPN 5 Ungaran tahun ajaran 2018-2019 dengan sample 60 dataset .
2. Pengklasifikasian penerima KIP menggunakan metode naive bayes berdasarkan faktor yang mempengaruhinya yaitu jenis kelamin, alamat, pekerjaan orang tua, penghasilan orang tua, pendidikan orang tua, dan penerima KPS (Kartu Perlindungan Sosial).

1.4 Tujuan Penelitian

Tujuan dari penelitian ini, penggunaan metode naïve bayes digunakan untuk mengetahui Kelayakan penerima KIP di SMPN 5 Ungaran dan mendapatkan akurasi data mining dengan metode naive bayes.

1.5 Manfaat Penelitian

1.5.1 Bagi SMPN 5 Ungaran

1. Mengetahui kategori penerima KIP di SMPN 5 Ungaran
2. Membantu pihak sekolah dalam mengetahui kriteria penerima KIP
3. Menambah wawasan dan referensi untuk mempelajari klasifikasi data mining menggunakan metode naïve bayes.

1.5.2 Bagi Universitas Dian Nuswantoro

1. Menambah kajian pustaka mengenai penelitian yang bersangkutan.
2. Menambah jurnal online dikampus yang telah ada.

1.5.3 Bagi Pembaca

1. Dapat mengimplementasikan klasifikasi data mining menggunakan metode naïve bayes dalam dataset siswa SMPN 5 Ungaran.
2. Menambah literatur bagi pembaca dalam menerapkan algoritma Naïve Bayes Classifier.

BAB II

TINJAUAN PUSTAKA

2.1 Tinjauan Studi

Hasil dari pencarian jurnal di Google Scholar, belum banyak penelitian yang mengangkat topik tentang pengklasifikasian Penerima KIP maka dari itu sebagai referensi dari penelitian yang di buat akan mengambil beberapa penelitian yang menggunakan Algoritma Naïve Bayes sebagai pengklasifikasian data masing-masing topik penelitian, di antaranya:

Dinda Ayu Muthia (2017) dalam penelitian ini digunakan metode pemilihan fitur, yaitu Genetic algorithm agar bisa meningkatkan akurasi pengklasifikasi Naïve Bayes. Penelitian ini menghasilkan klasifikasi teks dalam bentuk positif atau negatif dari review restoran. Pengukuran berdasarkan akurasi Naive Bayes sebelum dan sesudah penambahan metode pemilihan fitur. Evaluasi dilakukan menggunakan 10 fold cross validation. Sedangkan pengukuran akurasi diukur dengan confusion matrix dan kurva ROC. Hasil penelitian menunjukkan peningkatan akurasi Naïve Bayes dari 86.50% menjadi 90.50%.

Alfa Saleh (2015) menerapkan algoritma naïve bayes classifier untuk memprediksi besarnya penggunaan listrik tiap rumah tangga agar lebih mudah mengatur penggunaan listrik. dari 60 data penggunaan listrik rumah tangga yang diuji dengan metode naïve bayes, maka diperoleh hasil persentase 78,3333% untuk keakuratan prediksi, di mana dari 60 data penggunaan listrik rumah tangga yang diuji terdapat 47 data penggunaan listrik rumah tangga yang berhasil diklasifikasikan dengan benar.

Akhmad Pandhu Wijaya, Heru Agus Santoso (2016) Klasifikasi ini teknik dari data mining dan pertambangan teks juga digunakan untuk mencari atau mengatur kelas dibedakan dengan menggunakan beberapa fungsi dengan tujuan memungkinkan model untuk digunakan untuk data pengujian. Pada penelitian ini, objeknya adalah Situs Web Jawa Tengah dan diklasifikasikan oleh Naïve Bayes Classification (NBC). Dengan menggunakan metode ini diharapkan memfasilitasi klasifikasi dokumen bahasa Indonesia untuk identifikasi konten e-government.

Amrin, Hafdiarsya Saiyar (2018) Pada penelitian ini, menerapkan metode klasifikasi data mining, yaitu metode Naïve Bayes untuk mendiagnosa penyakit tuberculosis. Berdasarkan hasil pengukuran performa dari model tersebut dengan menggunakan metode pengujian Cross Validation, Confusion Matrix dan Kurva ROC, diketahui bahwa metode Naïve Bayes dengan tingkat akurasi sebesar 94,18% dan nilai area under the

curva (AUC) sebesar 0,977. Hal ini menunjukkan model yang dihasilkan termasuk kategori klasifikasi sangat baik karena memiliki nilai AUC antara 0.90-1.00.

Andriana Candra Dewi, Arief Andy Soebroto, M. Tanzil Furqon (2015) Dalam sistem pakar ini menggunakan metode Naive Bayes sebagai metode inferensi untuk mendiagnosis penyakit. Jenis penyakit yang dapat dikenali lewat sistem pakar ini adalah 11 jenis penyakit sedangkan gejala yang dapat dikenali sistem pakar adalah 20 jenis gejala. Hasil pengujian keakuratan 26 data kasus uji, telah menghasilkan tingkat persentase kesesuaian 96,15%.

Tabel 2.1 State of The Art

No .	Peneliti	Tahun	Judul	Masalah	Metode	Hasil
1.	Dinda Ayu Muthia	2017	Analisis Sentimen Pada Review Restoran Dengan Teks Bahasa Indonesia Menggunakan Algoritma Naïve Bayes	Seberapa besar efek metode pemilihan fitur <i>Genetic Algorithm</i> pada akurasi analisa sentiment pada review restoran dengan teks bahasa Indonesia	Naïve Bayes Classifier	Dari pengolahan data yang sudah dilakukan, Genetic Algorithm terbukti dapat meningkatkan akurasi pengklasifikasi Naïve Bayes. Data review restorandapat diklasifikasi dengan baik ke dalam bentuk positif dan negatif. Akurasi Naïve Bayes sebelum menggunakan penggabungan metode pemilihan fitur mencapai 86.50% Sedangkan setelah menggunakan penggabungan metode pemilihan fitur, yaitu Genetic Algorithm, akurasinya meningkat hingga mencapai 90.50%. Peningkatan akurasi mencapai 4%.

2.	Alfa Saleh	2015	Implementasi Metode Klasifikasi Naïve Bayes Dalam Memprediksi Basarnya Penggunaan Listrik Rumah Tangga	Penting nya listrik berdampak pada permintaan listrik yang semakin besar, tetapi tidak linier dengan persediaan listrik yang belum mampu memenuhi permintaan listrik yang begitu besar tersebut.	Naïve Bayes Classifier	Berdasarkan data rumah tangga yang dijadikan data training, metode Naive Bayes berhasil mengklasifikasikan 47 data dari 60 data yang diuji. Sehingga metode Naive Bayes berhasil memprediksi besarnya penggunaan listrik rumah tanggadengan persentase keakuratan sebesar 78,3333%.
----	------------	------	--	--	------------------------	---

3.	Akhmad Pandhu Wijaya & Heru Agus Santoso	2016	Naïve Bayes Classificat ion Pada Klasifikasi Dokumen Untuk Identifikasi Konten E- Governme Nt	Banyakny a informasi digital dalam bahasa Indonesia, perlu untuk clustering dokumen berdasarka n apa yang dicari sehingga untuk mendapatk an beberapa informasi dapat dilakukan dengan sesuai, ringkas, menyeluru h	Naïve Bayes	klasifikasi dokumen menggunakan Naïve Bayes Classifier pada penelitian ini dengan data training sebanyak 260 dokumen politik dan 222 dokumen ekonomi menggunakan 40 data testing menunjukkan nilai akurasi yang baik pada keseluruhan klasifikasi, dengan akurasi keseluruhan klasifikasi sebesar 85%.
----	---	------	---	---	----------------	---

4.	Amrin, Hafdiarsy a Saiyar	2018	Aplikasi Diagnosa Penyakit Tuberculosis Menggunakan algoritma Naïve Bayes	Penyakit Tuberculosis adalah penyakit yang mudah sekali menular . Sehingga sangat penting bagi untuk mendiagnosa secara dini penyakit ini untuk mengurangi penularan	Naïve Bayes Classifier	performa model naïve bayesmemberikan tingkat akurasi kebenaran sebesar 94,18% dengan nilai area under the curve (AUC) sebesar 0,977. Hal ini menunjukkan bahwa model tersebut termasuk katagori klasifikasi sangat baik karena memiliki nilai AUC antara 0.90-1.00.
----	---------------------------------	------	--	--	------------------------------	--

5.	Andriana Candra Dewi, Arief Andy Soebroto, M. Tanzil Furqon	2015	Sistem Pakar Diagnosa Penyakit Sapi Potong Dengan Metode Naïve Bayes	Dalam memproduksi daging sapi faktor penting yang harus dijaga adalah kesehatan sapi itu sendiri. Biasanya para peternak sapi tidak begitu kenal dengan berbagai penyakit sapi dan dokter pun tidak selalu tersedia	Naïve Bayes Classifier	pengujian akurasi terhadap variasi data menghasilkan nilai rata-rata akurasi masing- masing skenario sebesar 93,08%, 93,85%, 93,85%, 92,31% dan 92,31%. Sehingga didapatkan rata-rata akurasi sistem sebesar 93,08%. Tingkat akurasi tertinggi didapat ketika variasi data training berjumlah 40% dan 60% dari keseluruhan jumlah data training yang ada. Hal ini membuktikan bahwa komposisi data kasus berpengaruh dalam hasil akurasi sistem.
----	--	------	--	--	------------------------------	---

2.2 Tinjauan Pustaka

2.2.1 Klasifikasi

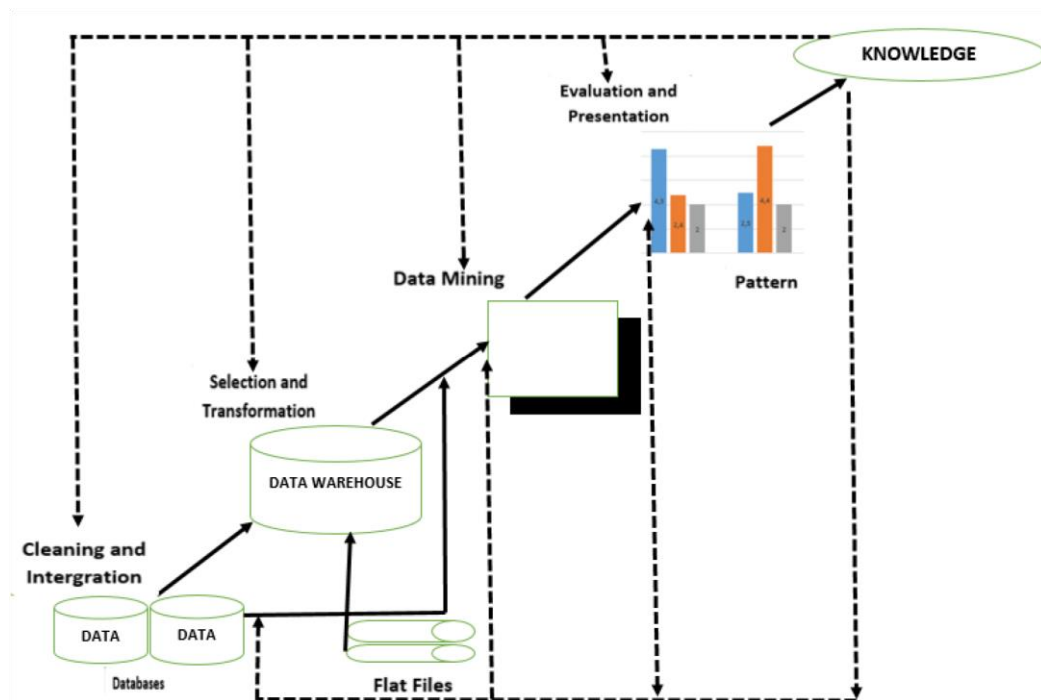
Program Indonesia Pintar Program Indonesia Pintar adalah salah satu program nasional yang bertujuan untuk meningkatkan angka partisipasi pendidikan dasar dan menengah, meningkatkan angka berkelanjutan pendidikan, menurunkan kesenjangan partisipasi pendidikan, meningkatkan kesiapan siswa pendidikan menengah untuk menghadapi pasar kerja. Program Indonesia Pintar diselenggarakan oleh Kementerian Pendidikan dan Kebudayaan dan Kementerian Sosial. Program Indonesia Pintar merupakan penyempurnaan dari program Bantuan Siswa Miskin (BSM) yang memberikan bantuan tunai pendidikan bagi anak usia sekolah yang berhak terutama dari keluarga pemegang KKS dan ditandai dengan Kartu Indonesia Pintar (KIP) TNP2K (2016). Selain KKS dan KIP, pihak sekolah juga bisa mendaftarkan siswanya yang kurang mampu dengan meminta keluarga siswa membawa surat keterangan PKH, KIS atau SKTM. SKTM adalah surat keterangan tidak mampu yang ditujukan untuk keluarga miskin dan bisa didapat melalui kelurahan setempat Pemerintah Kabupaten Bangil (2016). KKS atau Kartu Keluarga Sejahtera adalah bantuan non tunai melalui pembukaan rekening simpanan bagi rakyat kurang mampu Azarine (2016). PKH adalah Program Keluarga Sejahtera yang merupakan bantuan non tunai bagi Rumah Tangga Sangat Miskin (RTSM) yang akan menerima bantuan jika keluarga tersebut menyekolahkan anaknya dengan ingkat kehadiran tertentu Kementerian Sosial (2016). KPS adalah Kartu Perlindungan Sosial yang merupakan kartu untuk mendapatkan manfaat dari Program Subsidi Beras, Bantuan Siswa Miskin dan Bantuan Langsung Sementara Masyarakat TNP2K (2016). KIS atau Kartu Indonesia Sehat adalah penanda masyarakat miskin agar mendapatkan pelayanan kesehatan melalui Jaminan Kesehatan Nasional TNP2K (2016).

2.2.2 Data Mining

Pada dasarnya data mining tersusun atas algoritma atau metode yang memungkinkan seseorang untuk mendapatkan sebuah wawasan dasar dan pengetahuan pada suatu data yang besar. Data mining juga dapat dikatakan sebagai penggabungan bidang sistem database, statistik, pembelajaran mesin, dan pengenalan pola. Dalam pemrosesannya adapun tahapan pra-pengolahan seperti ekstraksi data, pembersihan data, fusi data, reduksi data dan fitur konstruksi.

2.2.2.1 Tahap-Tahap Data Mining

Berikut ini merupakan tahapan data mining yang sangat berpengaruh pada kualitas data :



Gambar 2.1 Tahap-tahap Data Mining

1. Pembersihan data (data cleaning)

Tahap pembersihan data yaitu proses menghilangkan noise dan data yang tidak konsisten atau tidak relevan. Data yang diperoleh, baik dari database suatu perusahaan maupun hasil eksperimen, memiliki isian-isian yang tidak sempurna seperti data yang hilang, data yang tidak valid atau juga hanya sekedar salah ketik. Selain itu, ada juga atribut-atribut data yang tidak relevan yang lebih baik

dibuang. Pembersihan data akan mempengaruhi performansi dari teknik data mining karena data yang ditangani akan berkurang jumlah dan kompleksitasnya.

2. Integrasi data (data integration)

Tahap integrasi data merupakan penggabungan data dari berbagai database ke dalam satu database baru. Integrasi data dilakukan pada atribut-atribut yang mengidentifikasi entitas-entitas yang unik seperti atribut usia, jenis pidana umum, nomor identitas dan lainnya. Hal ini dilakukan secara cermat karena kesalahan pada integrasi data bisa menghasilkan hasil yang menyimpang dan bahkan menyesatkan pengambilan aksi nantinya. Sebagai contoh bila integrasi data berdasarkan jenis pidana umum ternyata menggabungkan data pidana umum dari kategori yang berbeda maka akan didapatkan korelasi antar data pidana umum yang sebenarnya tidak ada.

3. Seleksi Data (Data Selection)

Tahap seleksi data yang terdapat pada database sering kali tidak semuanya dipakai, maka hanya data yang sesuai untuk dianalisis dan akan diambil dari database. Sebagai contoh, sebuah kasus yang meneliti pola kecenderungan konsumen dalam membeli kebutuhan pada sebuah supermarket, hasil analisis tidak perlu mengambil nama pelanggan, cukup dengan id pelanggan saja.

4. Transformasi data (Data Transformation)

Tahap transformasi data yaitu data diubah atau digabung ke dalam format yang sesuai untuk diproses dalam data mining. Beberapa metode data mining membutuhkan format data yang khusus sebelum diaplikasikan. Sebagai contoh beberapa metode standar seperti analisis asosiasi dan clustering hanya bisa menerima input data kategorikal. Karenanya data berupa angka numerik yang berlanjut perlu dibagi-bagi menjadi beberapa interval. Proses ini sering disebut transformasi data.

5. Proses mining,

Proses mining ini adalah proses yang utama saat metode diterapkan untuk menemukan sebuah pengetahuan berharga dan tersembunyi dari data.

6. Evaluasi pola (pattern evaluation),

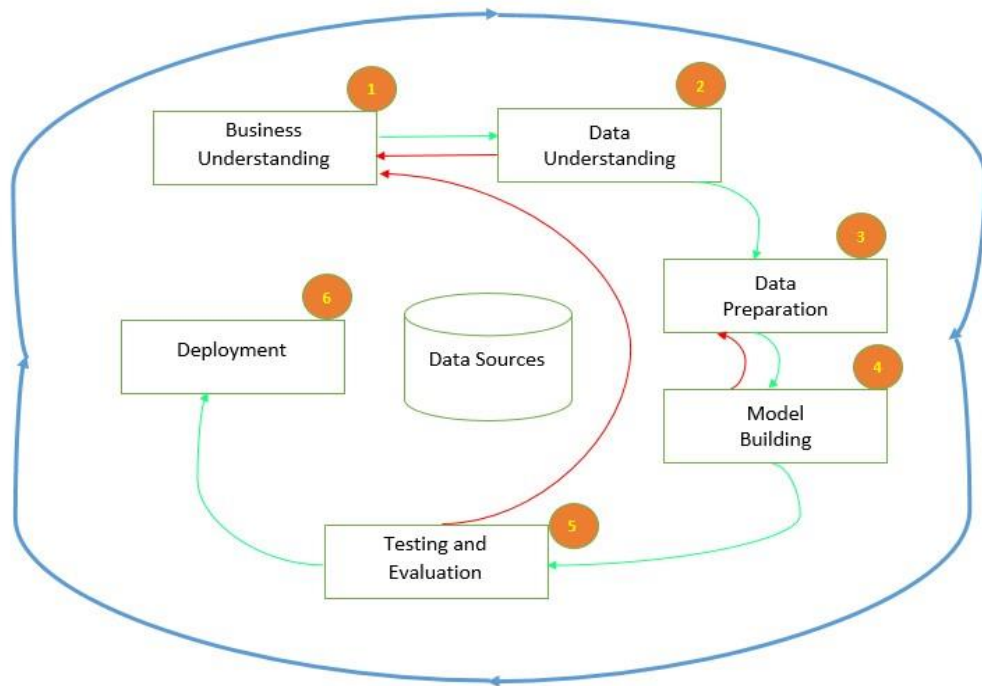
Tahap evaluasi pola dapat digunakan untuk mengidentifikasi pola-pola menarik kedalam knowledge based yang ditemukan. Dalam tahap ini hasil dari teknik data mining berupa pola-pola yang khas maupun model prediksi dievaluasi untuk menilai apakah hipotesa yang ada memang tercapai. Bila ternyata hasil yang diperoleh tidak sesuai hipotesa ada beberapa alternatif yang dapat diambil seperti menjadikannya umpan balik untuk memperbaiki proses data mining, mencoba metode data mining lain yang lebih sesuai, atau menerima hasil ini sebagai suatu hasil yang di luar dugaan yang mungkin bermanfaat.

7. Presentasi pengetahuan (knowledge presentation),

Presentasi pengetahuan merupakan proses visualisasi dan penyajian pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna. Tahap terakhir dari proses data mining adalah bagaimana memformulasikan keputusan atau aksi dari hasil analisis yang didapat agar dipahami oleh orang tidak mengetahui data mining.

2.2.2.2 CRISP-DM

Cross-Industry Standard Process Model for Data Mining menjelaskan tentang proses data mining dalam enam tahap. Salah satu tujuan dari proses ini untuk menemukan pola yang menarik dan bermakna dalam data. Keuntungan dari menggunakan proses ini adalah menjelaskan langkah paling umum dalam setiap prosesnya. Selain itu, proses ini juga melibatkan manajer dan praktisi secara bersamaan. Dimana manajer memberikan arahan tujuan utama proyek yang akan dikerjakan, ketersediaan data dan model - model yang akan digunakan. Sedangkan praktisi bekerja sesuai dengan bidangnya dalam setiap proses yang ada, praktisi disinipun bisa terdiri dari berbagai disiplin ilmu, bisa dari matematika, statistika maupun teknik informatika.



Gambar 2.2 Crisp-Data Mining

1. Business Understanding

Tahap yang pertama dalam CRISP-DM dan termasuk bagian yang cukup vital dalam proyek dan selama pengerjaan pada proses selanjutnya. Dibutuhkan pengetahuan dari objek bisnis, bagaimana membangun atau mendapatkan data, dan bagaimana untuk mencocokkan tujuan pemodelan untuk tujuan bisnis sehingga model terbaik dapat dibangun.

2. Data Understanding

Pada tahapan ini diharuskan untuk dapat mengidentifikasi masalah dalam data. Selain itu, memberikan analisi dasar untuk sebuah proyek dengan membuat ringkasan dan mengidentifikasi potensi masalah dalam data. Tahap ini juga harus dilakukan secara cermat dan tidak terburu - buru, seperti pada visualisasi data yang terkadang sangat sulit didapat jika dihubungkan dengan ringkasan datanya. Apabila ada masalah pada tahap ini yang belum terjawab, maka akan mengganggu pada tahap modeling. Ringkasan dari data berguna untuk mengkonfirmasi apakah data terdistribusi seperti yang diharapkan, atau mengungkapkan penyimpangan tak terduga yang perlu ditangani pada tahap selanjutnya, yaitu Data Preparation. Masalah dalam data biasanya seperti nilai-nilai yang hilang, outlier, berdistribusi spike, berdistribusi bimodal harus diidentifikasi dan diukur sehingga dapat diperbaiki dalam Data Preparation.

3. Data Preparation

Tahap ini membutuhkan pemikiran yang cukup matang dan usaha yang cukup tinggi untuk memastikan data yang diperoleh sesuai dengan algoritma yang digunakan. Bukan berarti pada tahap ini pertama kali dimana masalah-masalah pada data sudah diselesaikan, namun data sudah dapat digunakan hingga tahap terakhir. Tahap ini sering ditinjau kembali saat menemukan masalah pada saat pembangunan model. Sehingga dilakukan iterasi sampai menemukan hal yang cocok dengan data. Tahap sampling dapat dilakukan disini dan data secara umum dibagi menjadi dua, data training dan data testing.

4. Modeling

Tahap ini akhirnya kita dapat menggunakan Statistika dan Machine Learning untuk mendapatkan pengetahuan yang berguna dari data untuk mencapai tujuan proyek. Beberapa modeling yang biasa dilakukan adalah classification, scoring, ranking, clustering, finding relation, characterization.

5. Evaluation

Setelah mempunyai model, kita harus menentukan apakah sesuai dengan tujuan kita. Pada tahap ini ada beberapa pertanyaan yang bisa membantu apakah model kita sudah sesuai dengan tujuan atau belum, apakah cukup akurat untuk kebutuhan kita dan hasil sesuai perkiraan dan lainnya.

6. Deployment

Tahap ini adalah tahap yang paling dihargai dari proses CRISP-DM. Perencanaan untuk deployment dimulai selama Business Understanding dan harus menggabungkan tidak hanya bagaimana untuk menghasilkan nilai model, tetapi juga bagaimana mengkonversi skor keputusan, dan bagaimana untuk menggabungkan keputusan dalam sistem operasional. Pada akhirnya, rencana sistem Deployment mengakui bahwa tidak ada model yang statis. Model tersebut dibangun dari data yang diwakili data pada waktu tertentu, sehingga perubahan waktu dapat menyebabkan berubahnya karakteristik data. Modelpun harus dipantau dan mungkin diganti dengan model yang sudah diperbaiki.

2.2.3 Klasifikasi

Dalam buku *Data Mining and Analysis: Fundamental Concepts and Algorithms*, Tugas klasifikasi untuk memprediksi label atau kelas untuk titik berlabel diberikan. [4] Klasifikasi adalah model atau fungsi M yang memprediksi label kelas y untuk diberikan masukan contoh x , yaitu $y = M(x)$, di mana $y \in \{c_1, c_2, \dots, c_k\}$ dan setiap c_i adalah label kelas (Nilai atribut kategorikal). Untuk membangun model yang membutuhkan satu set poin dengan label kelas yang benar, yang disebut training set. Setelah mempelajari model M , maka otomatis memprediksi kelas untuk setiap titik baru. Berbagai jenis klasifikasi model telah diusulkan seperti pohon keputusan, pengklasifikasi probabilistik, dukungan mesin vektor, dan sebagainya.

2.2.4 Naïve Bayes Classifier

Pelatihan dataset D terdiri dari n poin x_i dalam ruang d -dimensi, dan biarkan y_i menunjukkan kelas untuk setiap titik, dengan $y_i \in \{c_1, c_2, \dots, c_k\}$. Bayes classifier akan menggunakan teoremanya untuk memprediksi kelas untuk contoh uji baru, x . Dengan memperkirakan posterior probabilitas $P(c_i | x)$ untuk setiap c_i kelas, dan memilih kelas yang memiliki kemungkinan terbesar. kelas diperkirakan untuk x :

$$\hat{y} = \underset{c_i}{\operatorname{argmax}}\{P(c_i|x)\}$$

Teorema Bayes memungkinkan kita untuk membalikkan probabilitas posterior dalam hal kemungkinan dan probabilitas sebelumnya, sebagai berikut:

$$P(c_i|x) = \frac{P(x|c_i) \cdot P(c_i)}{P(x)}$$

dimana $P(x | c_i)$ adalah kemungkinan, didefinisikan sebagai probabilitas mengamati x asumsi bahwa kelas yang benar adalah c_i , $P(c_i)$ adalah probabilitas sebelum kelas c_i , dan $P(x)$ adalah probabilitas mengamati x dari salah satu kelas k , diberikan sebagai

$$P(x) = \sum_{j=1}^k P(x|c_j) \cdot P(c_j)$$

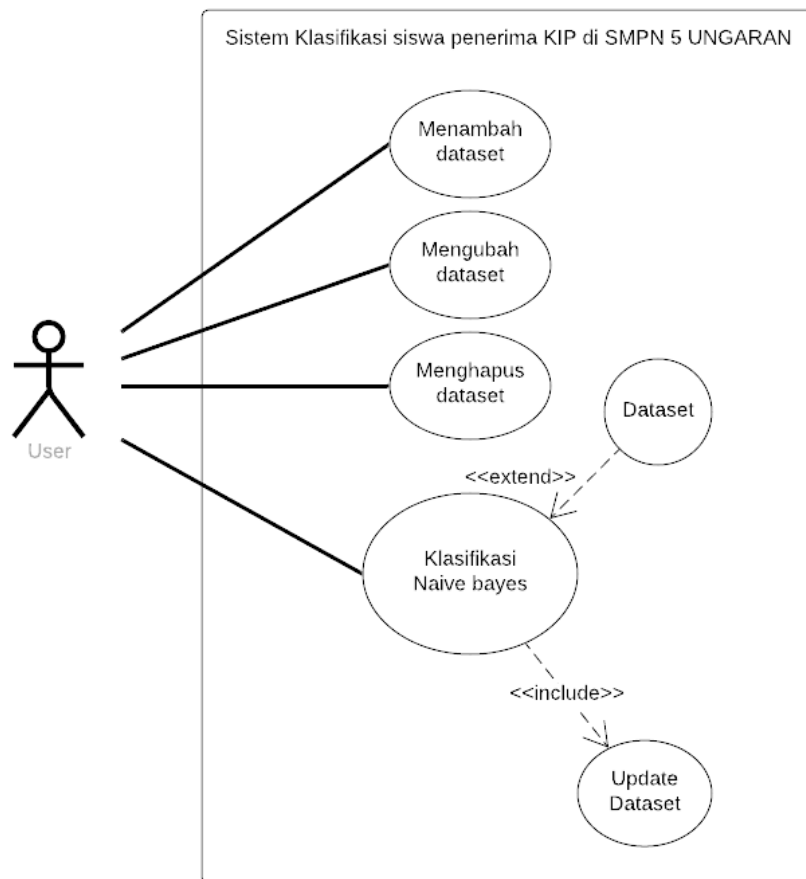
karena $P(x)$ adalah tetap untuk suatu titik tertentu, aturan Bayes dapat ditulis kembali sebagai

$$\hat{y} = \underset{c_i}{\operatorname{argmax}}\{P(c_i|x)\}$$

$$\hat{y} = \underset{c_i}{\operatorname{argmax}} P(c_i|x) = \frac{P(x|c_i) \cdot P(c_i)}{P(x)} = \underset{c_i}{\operatorname{argmax}}\{P(c_i|x)P(c_i)\}$$

dengan kata lain, kelas diprediksi dasarnya tergantung pada kemungkinan kelas yang mengambil probabilitas sebelum ke dalam laporan.

2.2.5 UML



Gambar 2.3 Use Case Diagram Sistem

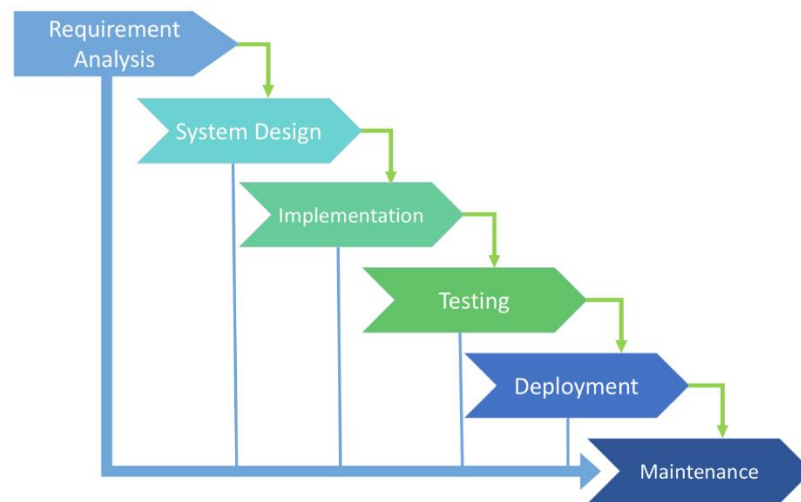
Dalam sebuah proses klasifikasi penerima KIP dibutuhkan seorang Operator atau Admin yang bertugas mengatur jalannya sistem. Operator ini dapat melakukan Tambah atau Input data, Simpan, Ubah dan Hapus .

Untuk proses sistem klasifikasi metode naive bayes terjadi saat seorang Operator mengubah data Penerima KIP, maka sistem klasifikasi juga akan berubah. Misal : semula penerima KIP tidak memiliki kartu KPS maka akan terklasifikasi tidak layak untuk mendapatkan kartu KIP. Setelah itu, data dapat di simpan kembali atau dihapus sesuai kebutuhan.

2.2.6 Model Pengembangan Sistem

Waterfall

Dalam pengembangannya metode waterfall memiliki beberapa tahapan yang runtut: requirement (analisis kebutuhan), design sistem (system design), Coding & Testing, Penerapan Program, pemeliharaan.



Gambar 2.4 Metode Waterfall

- Requirement (analisa kebutuhan)

Dalam tahapan ini melakukan analisa terhadap kebutuhan sistem. Pengumpulan data dapat dilakukan dengan penelitian, wawancara atau study literatur. Seorang analisa akan menggali informasi sebanyak-banyaknya dari user sehingga akan tercipta sebuah sistem komputer yang bisa melakukan tugas-tugas yang diinginkan oleh user tersebut. Sehingga tahap ini menghasilkan dokumen user requirement atau bisa dikatakan sebagai data yang berhubungan dengan keinginan user dalam pembuatan sistem. Dokumen inilah yang akan menjadi acuan sistem analisis untuk menterjemahkan kedalam bahasa pemrograman.

- Design System (design sistem)

Proses design akan menterjemahkan syarat kebutuhan sebuah perancangan perangkat lunak yang dapat diperkirakan sebelum dibuat koding. Proses ini berfokus pada : struktur data, arsitektur perangkat lunak, representasi interface, dan detail (algoritma) prosedural. Tahapan ini akan menghasilkan dokumen yang disebut software requirement yang kemudian akan digunakan programmer untuk melakukan aktivitas pembuatan sistemnya.

- Coding & Testing (penulisan kode program / implementation)

Coding merupakan penerjemahan design dalam bahasa yang bisa dikenali oleh komputer. Dilakukan oleh programmer untuk menterjemahkan transaksi yang diminta oleh user. Penggunaan computer akan dimaksimalkan dalam tahapan ini. Setelah pengkodean selesai maka akan dilakukan testing atau percobaan terhadap sistem yang telah dibuat dengan tujuan menemukan kesalahan-kesalahan terhadap system tersebut dan kemudian bisa diperbaiki.

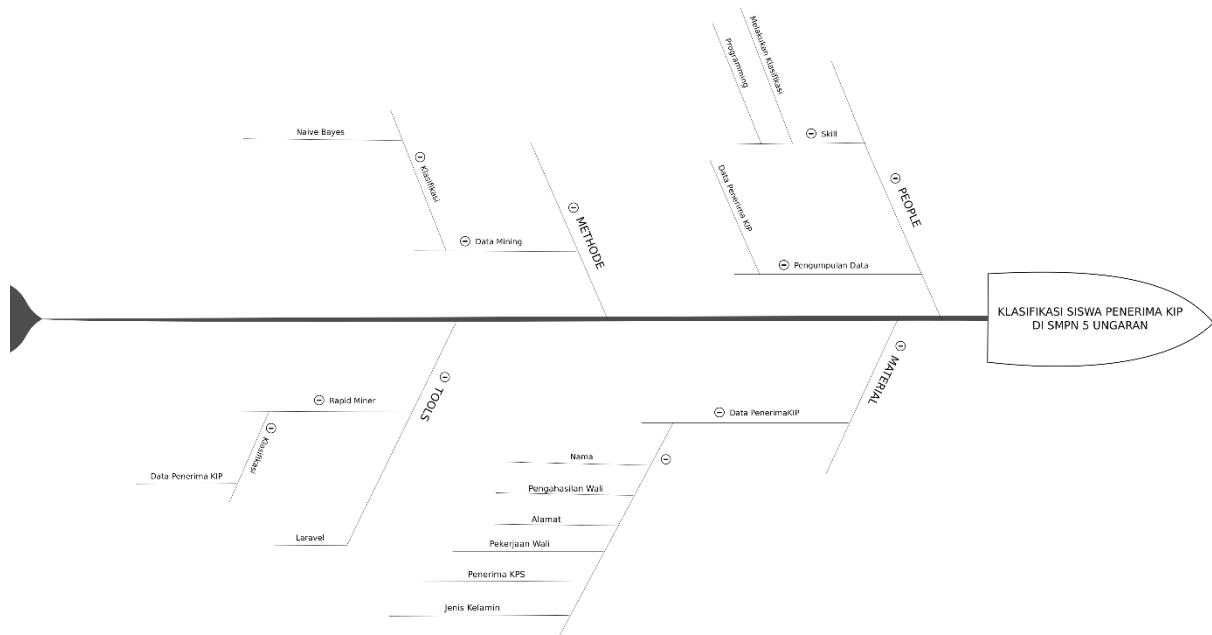
- Penerapan / Pengujian Program (Integration & Testing)

Tahapan ini bisa dikatakan final dalam pembuatan sebuah sistem. Setelah melakukan analisa, design dan pengkodean maka sistem yang sudah dapat digunakan oleh user.

- Pemeliharaan (Operation & Maintenance)

Perangkat lunak yang susah disampaikan kepada pelanggan pasti akan mengalami perubahan. Perubahan tersebut bisa karena mengalami kesalahan karena perangkat lunak harus menyesuaikan dengan lingkungan (periperal atau sistem operasi baru) baru, atau dengan alasan user membutuhkan perkembangan fungsional.

2.3 Kerangka Pemikiran



Gambar 2.5 Fishbone Diagram

Dari kerangka diatas, dapat dijelaskan bahwa masalah yang timbul yaitu tentang Klasifikasi Penerima KIP. Masalah ini bermula dari banyaknya kriteria yang ditentukan untuk dapat menerima KIP . yang sebelumnya dikelompokkan menjadi 2 jenis ya atau tidak layak dengan faktor yang mempengaruhinya yaitu Jenis kelamin, alamat, pendapatan orang tua , penghasilan orang tua, pendidikan orang tua, penerima KPS. Proses pengelompokan ini dinamakan klasifikasi yang akan menggunakan metode naive bayes pada data mining dengan tools RapidMiner5 dan aplikasi berbasis web dengan framework laravel.

BAB III

MEDOTE PENELITIAN

3.1 Metode Pengumpulan Data

Data dalam penelitian ini menggunakan data yang diperoleh melalui tata usaha SMPN 5 Ungaran . Selain itu, juga digunakan beberapa Studi Pustaka (*Library Research Method*) antara lain buku yang membahas data mining, jurnal mengenai penggunaan metode naive bayes dan *E-Book* mengenai data mining beserta algoritmanya.

3.2 Teknik Analisa Data

Data awal jumlah record atau data awal yang diperoleh dari pengumpulan data sebanyak 60 dataset dengan 6 atribut diantaranya Penerima KPS yang mempengaruhi sebuah label yaitu layak tidak nya menerima KIP. Dalam penelitian ini setelah data didapatkan ada beberapa tahapan yang dilakukan terhadap data dengan beberapa tahap pengolahan awal data (*preparation data*) untuk mendapatkan data yang berkualitas, beberapa teknik yang dilakukan sebagai berikut :

1. Data Cleaning

Untuk mendapatkan data yang berkualitas perlu adanya identifikasi atau menghilangkan *outliner* dan menghilangkan data *nois*. Kemudian mengisi nilai-nilai yang hilang atau data yang tidak lengkap (*missing value*). Dalam hal ini algoritma naive bayes mempunyai keunggulan tersendiri dalam menangani data yang tidak lengkap tersebut.

2. Data Integration and transformation

Kemudian data akan melalui tahap *integration* yang digunakan dalam menganalisis data korelasi, atribut yang redundan dan duplikat data dan *transformation* digunakan untuk meningkatkan akurasi dan efisiensi algoritma.

3. Data reduction

Mereduksi data set dengan mengurangi jumlah atribut dan record agar lebih sedikit namun tetap bersifat informative. Berikut atribut data awal yang belum dilakukan reduction data :

Tabel 3.1 Detail Atribut data penelitian

No.	Atribut
1.	No
2.	Nama
3.	Jenis Kelamin
4.	Alamat
5.	Penghasilan orang tua
6.	Pekerjaan orang tua
7.	Pendidikan orang tua
8.	Penerima KPS

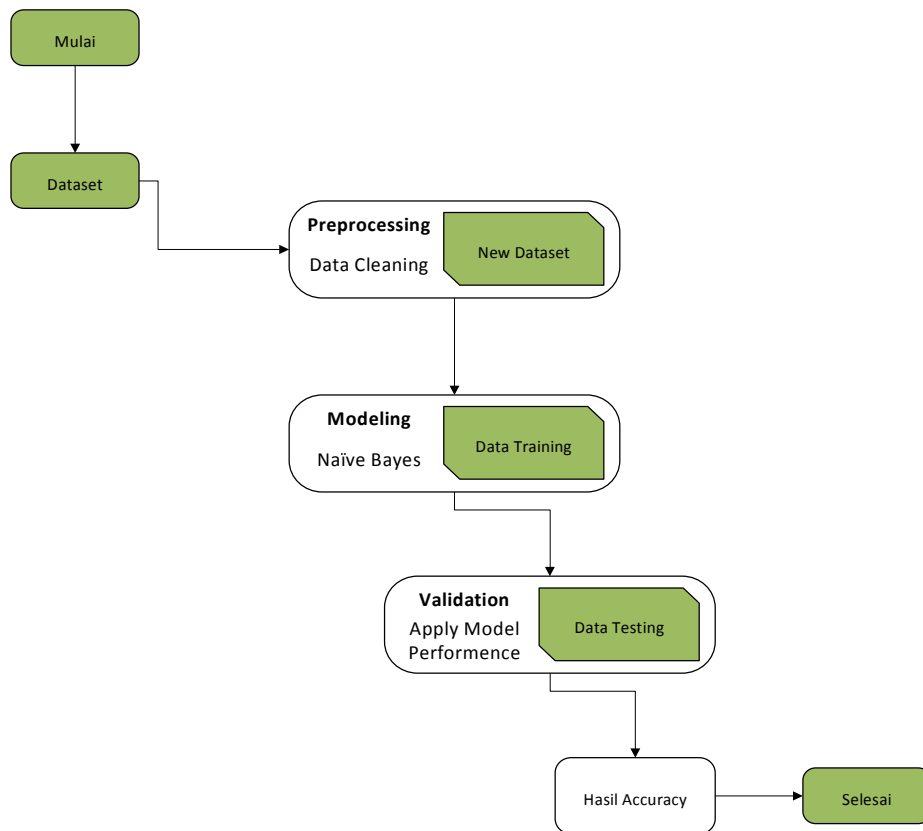
Data yang telah dilakukan tahapan pengolahan (*preparation data*) menghasilkan atribut sebagai berikut :

Tabel 3.2 Atribut data yang digunakan

No.	Atribut
1.	Jenis Kelamin
2.	Alamat
3.	Penghasilan orang tua
4.	Pekerjaan orang tua
5.	Pendidikan orang tua
6.	Penerima KPS

3.3 Model Yang Diusulkan

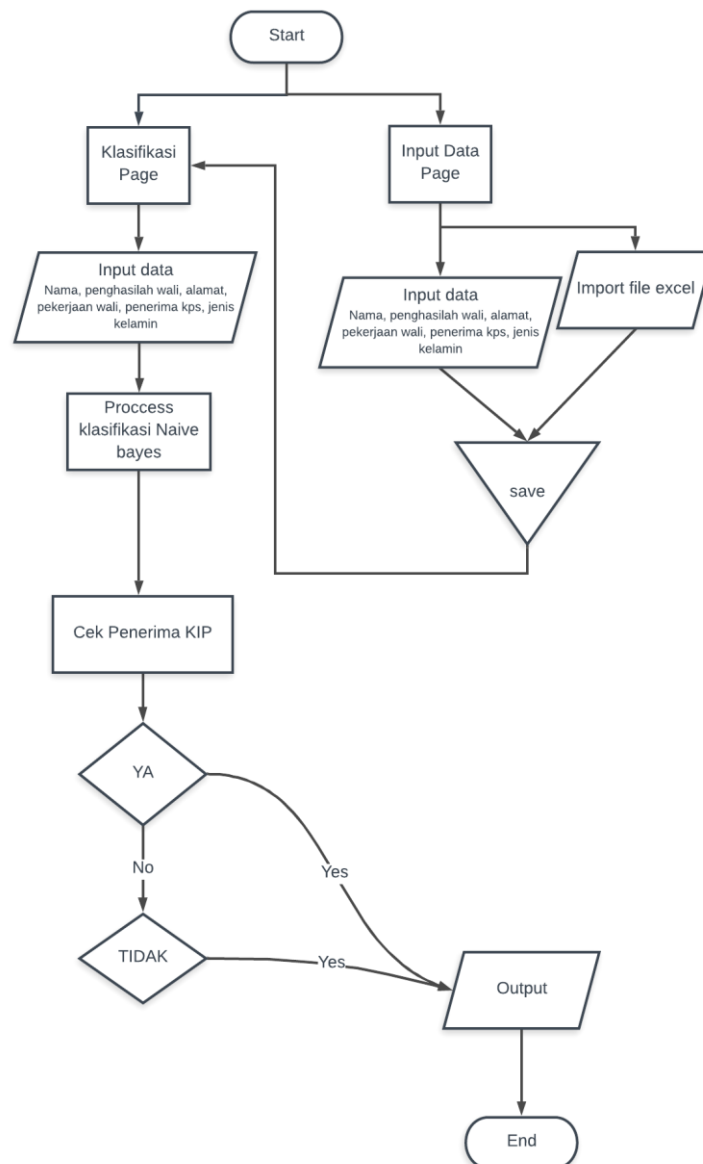
3.3.1 Modelling



Gambar 3.1 Flowchart untuk menemukan
Performance Algoritma

Metode yang akan digunakan dalam penelitian ini adalah Algoritma Naive Bayes yang akan dicari performance nya yaitu Performance Vektor (accuracy) dan Confusion Matrix. Data yang digunakan berupa data yang telah diproses data clean, dan untuk melakukan pengukurannya menggunakan tools RapidMiner5.

3.3.2 Naïve Bayes Classifier



Algortima Naive Bayes

Dimulai dari Input data siswa pada SMPN 5 Ungaran berupa Jenis kelamin, Alamat, Penghasilan Orang tua, Pekerjaan Orang tua, Pendidikan Orang tua, Penerima KPS Kemudian, penerima akan di klasifikasikan menggunakan metode naive bayes layak atau tidak nya untuk menerima KIP.

Apabila penerima tidak sesuai dengan kriteria penerima KIP maka akan di identifikasi kembali menuju peninjauan data Penerima KIP. Apabila penerima telah

cocok dengan salah satu kriteria maka akan muncul hasil berupa layak nya untuk menerima KIP.

3.4 Pengujian Metode Naive Bayes

1. Menghitung $P(C_i)$ untuk setiap kelas: Nilai $P(C_i)$ untuk i

Menghitung peluang kelas (i) :

$$P(class(i)) = \text{jumlah kelas (i)} / \text{jumlah keseluruhan data.}$$

Menghitung peluang pada masing-masing atribut :

$$P(X|C_i) = \text{jumlah } x / \text{jumlah kelas (i).}$$

2. Menghitung $P(X|C_i)$ untuk semua kelas dengan mengalikan semua variabel.

$$\sum P(X|C_i) * P(C_i)$$

3. Membandingkan hasil perkalian $P(X|C_i)$,setelah itu mengambil hasil yang terbesar sebagai hasilnya.

BAB IV

JADWAL PENELITIAN DATA MINING

4.1 Jadwal Penyusunan Penelitian Data Mining

Dalam melakukan penelitian Data Mining, kami melakukan tahapan dengan jadwal penyusunan sebagai berikut:

Tabel 4.1 Jadwal Penyusunan Penelitian Data Mining

No	Kegiatan	Maret				April				Mei				Juni			
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	Pengumpulan data																
2	Membuat Bab I																
3	Membuat Bab II																
4	Membuat Bab III																
5	Mencari sumber pustaka																
6	Sorting & Cleaning data																
7	Mengolah data																
8	Membuat Aplikasi																
9	Membuat Bab IV																
10	Membuat Bab V																

DAFTAR PUSTAKA

- [1] <https://www.cermati.com/artikel/kartu-indonesia-pintar-apa-itu-manfaat-dan-siapa-berhak-menerima> diakses 11 April 2019
- [2] E. U. A. S. Bonifacius Vicky Indriyono, “Klasifikasi Jenis Buku Berdasarkan Judul Dan Sinopsis Menggunakan Metode Naïve Bayes Classifier (Studi Kasus: STMIK KADIRI),” *J. Sist. Inf.*, vol. 5, no. 4, 2015.
- [3] M. Ridwan, H. Suyono, and M. Sarosa, “Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier,” *Eeccis*, vol. 7, no. 1, pp. 59–64, 2013.
- [4] M. J. Zaki and M. J. Meira, *Data Mining and Analysis: Fundamental Concepts and Algorithms*. 2014.
- [5] Dinda Ayu Muthia, “Analisis Sentimen Pada Review Restoran Dengan Teks Bahasa Indonesia Menggunakan Algoritma Naïve Bayes,” Vol.2. No. 2 Februari 2017
E-ISSN: 2527-4864
- [6] Alfa Saleh, “Implementasi Metode Klasifikasi Naïve Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga” *Citec Journal*, Vol. 2, No. 3, Mei 2015 –Juli 2015 ISSN: 2354-5771
- [7] A. P. Wijaya and H. A. Santoso, “Naïve Bayes Classification Pada Klasifikasi Dokumen Untuk Identifikasi Konten E-Government,” *J. Appl. Intell. Syst.*, vol. 1, no. 1, pp. 48–55, 2016.
- [8] Amrin, Hafdiarsya saiyar “Aplikasi Diagnosa Penyakit Tuberculosis Menggunakan Algoritma Naïve Bayes” *Jurnal Riset Komputer (JURIKOM)*, Vol. 5 No. 5, Oktober 2018 ISSN 2407-389X (Media Cetak) Hal: 498-502
- [9] Andriana Candra Dewi, Arief Andy Soebroto, M. Tanzil Furqon, “Sistem Pakar Diagnosa Penyakit Sapi Potong Dengan Metode Naïve Bayes” *Journal of Environmental Engineering & Sustainable Technology JEEST* Vol. 02 No. 02, November 2015, Pages 72-78
- [10] Nofriansyah, Dicky. 2014. *Konsep Data Mining Vs Sistem Pendukung Keputusan*. Yogyakarta: Deepublish.

- [11] <http://mbahsecond.blogspot.co.id/2013/10/metodologi-pengembanganwaterfall.html>
diakses pada 1 April 2019
- [12] <http://ariplie.blogspot.co.id/2015/04/pengertian-dan-tahap-tahap-datamining.html>
diakses tanggal 4 April 2019.
- [13] <http://nur-statistika-unpad.blogspot.co.id/2015/07/crisp-data-mining.html> diakses
tanggal 5 April 2019.
- [14] <http://yusnitadewi.ilearning.me/2013/10/01/eksplorasi-data-mining-software-rapid-miner/> diakses tanggal 7 April 2019.