



PENERAPAN ALGORITMA NAIVE BAYES UNTUK MEMPREDIKSI PENERIMAAN CALON ATLET PUSAT PENDIDIKAN DAN LATIHAN PELAJAR CABANG OLAHRAGA PENCAK SILAT (STUDI KASUS IPSI KABUPATEN BEKASI)

Nurhadi Surojudin¹, Abdul Halim Anshor²

Program Studi Teknik Informatika Fakultas Teknik Universitas Pelita Bangsa
¹nurhadi@pelitabangsa.ac.id, ²abdulhalimanshor@pelitabangsa.ac.id

Abstrak

Pencak silat adalah olahraga bela diri yang telah ada di Indonesia sejak lama dan terpelihara hingga saat ini. Prestasi adalah tujuan akhir dari proses pelatihan yang dijalani atlet olahraga pencak silat. Pemilihan atlet yang berkompentensi merupakan awal dari proses menciptakan atlet yang berprestasi. Setiap kelompok atau organisasi melakukan seleksi terhadap calon atlet yang akan dibina menjadi atlet yang menghasilkan prestasi. Seleksi dapat dilakukan dengan menggunakan pendukung keputusan, maka di perlukan adanya penilaian yang akurat salah satunya menggunakan *Algoritma*. *Algoritma* adalah suatu urutan atau alur yang dipakai dalam perhitungan atau pemecahan masalah.

Pada penelitian ini data yang digunakan adalah data calon atlet cabang olahraga pencak silat kabuapten bekasi masih berupa raw data, yang berisi atribut nama, tempat tanggal lahir, jenis kelamin, usia, berat badan (Kg), tinggi badan (Cm), nilai IMT, massa otot, massa lemak, dan keterangan. Kemudian data yang akan digali yaitu nilai dari setiap kriteria tersebut menjadi patokan untuk menilai kelayakan calon atlet. Dengan mengamati persoalan diatas data mining dengan metode naive bayes sangat tepat digunakan untuk menghasilkan Knowledge kelompok prioritas diantara ratusan data calon atlet pencak silat studi kasus IPSI. Adapun data yang akan dijadikan sebagai data training yaitu sebanyak 700 record data, data training digunakan untuk melakukan proses pelatihan algoritma terhadap dataset untuk menghasilkan suatu klasifikasi. Sedangkan data testing digunakan untuk melakukan pengujian terhadap proses klasifikasi yang digunakan algoritma terhadap dataset.

Data mining, Algoritma C4.5, Pencak Silat, Atlit

Abstract

Pencak silat is a martial art that has existed in Indonesia for a long time and has been preserved to this day. Achievement is the ultimate goal of the training process experienced by pencak silat athletes. Selection of competent athletes is the beginning of the process of creating outstanding athletes. Each group or organization selects prospective athletes who will be coached to become athletes who produce achievements. Selection can be done using decision support, so an accurate assessment is needed, one of which is using an algorithm. Algorithm is a sequence or flow that is used in calculations or problem solving.

In this study, the data used were data on prospective athletes for the pencak silat sport in Bekasi Regency, still in the form of raw data, which contained the attributes of name, place of birth date, gender, age, weight (Kg), height (Cm), BMI value, muscle mass, fat mass, and description. Then the data to be explored,

namely the value of each of these criteria, becomes the benchmark for assessing the eligibility of prospective athletes. By observing the problems above, data mining using the Naive Bayes method is very appropriate to use to generate knowledge of priority groups among hundreds of data on prospective pencak silat athletes from IPSI case studies. The data that will be used as training data is as much as 700 data records, the training data is used to carry out the algorithm training process on the dataset to produce a classification. While data testing is used to test the classification process used by the algorithm on the dataset.

Data mining, Algorithm C4.5, Pencak Silat, Atli

1. Pendahuluan

Pencak silat adalah olahraga bela diri yang telah ada di Indonesia sejak lama dan terpelihara hingga saat ini. Prestasi adalah tujuan akhir dari proses pelatihan yang dijalani atlet olahraga pencak silat.

Pemilihan atlet yang berkompentensi merupakan awal dari proses menciptakan atlet yang berprestasi. Setiap kelompok atau organisasi melakukan seleksi terhadap calon atlet yang akan dibina menjadi atlet yang menghasilkan prestasi. Penyeleksian atlet yang dibina memiliki kriteria-kriteria yang disesuaikan oleh masing-masing kelompok. Contohnya kemampuan fisik, ketahanan fisik, sikap, dan lain-lain. Kriteria penyeleksian harus disusun dengan baik agar proses pelatihan yang dilaksanakan nantinya dapat diikuti dengan baik dan menghasilkan atlet berprestasi.

Selama ini pelatih dalam penentuan calon atlet pencak silat hanya menggunakan perkiraan semata, akibatnya sering terjadi kesalahan dalam penerimaan calon atlet. Kondisi ini menyebabkan atlet sering tidak bisa maksimal dalam berlatih dan berdampak prestasi atlet tidak optimal dalam kompetisi. Kesalahan dalam pengambilan keputusan dapat membawa dampak yang sangat besar bagi prestasi atlet-atlet dan prestasi daerah itu sendiri. Hal ini merugikan para atlet yang mengikuti pertandingan.

Penentuan calon atlet pencak silat dengan cara melihat data dan menentukan siapa yang layak masuk dalam pelatihan pencak silat, karena dipengaruhi oleh faktor *human error* dengan penerimaan calon atlet yang dilakukan masih menggunakan cara manual. Sehingga mengakibatkan tingkat kesalahan sangat besar dan membutuhkan waktu yang lama. Oleh sebab itu, dibutuhkan solusi yang dapat membantu dalam pengambilan keputusan untuk mengatasi permasalahan tersebut. Data calon atlet pencak silat didapat dari data IPSI (Ikatan Pencak Silat Indonesia) yang berjumlah 300 data yang terdiri dari beberapa sekolah di daerah Kabupaten Bekasi. Data terdiri dari Usia, Berat badan, tinggi badan, massa otot dan massa lemak.

Selama ini belum ada metode yang mempermudah pelaksanaan seleksi penerimaan atlet pencak silat. Seleksi dapat dilakukan dengan menggunakan pendukung keputusan, maka di perlukan adanya penilaian yang akurat salah satunya menggunakan Algoritma. Algoritma adalah suatu urutan atau alur yang dipakai dalam perhitungan atau pemecahan masalah.

Keputusan adalah sistem yang membantu manusia dalam pengambilan keputusan. Pengambilan keputusan merupakan sesuatu yang sering dihadapi manusia, keputusan yang diambil biasanya karena faktor-faktor tertentu atas dasar logika dan lain-lain. Sistem pengambilan keputusan berbasis komputer membantu dalam pengambilan keputusan dengan menggunakan metode Algoritma Naive Bayes.

2. Landasan Pemikiran

2.1. Tinjauan Pustaka

“Penerapan Klasifikasi Bayes Untuk Memprediksi Jenis Latihan Siswa Pencak Silat (Studi Kasus Pencak Silat PSHT)” [1] dalam penelitian ini berisi tentang metode *naive bayes* untuk klasifikasi jenis latihan siswa pencak silat. Selama ini pelatih dalam penentuan jenis latihan pencak silat hanya menggunakan perkiraan intuisi semata, akibatnya sering terjadi ketidaksesuaian antara calon siswa dan jenis latihan yang diikuti. Kondisi ini menyebabkan sering

tidak bisa maksimal dalam berlatih disamping berdampak prestasi siswa tidak optimal dalam kompetisi.

Dalam penelitian ini dikembangkan aplikasi untuk memprediksi jenis latihan siswa berdasarkan karakteristik yang sesuai menggunakan klasifikasi Naive Bayes. Aplikasi ini diharapkan dapat membantu pelatih dalam meningkatkan kemampuan siswa sesuai karakteristiknya sehingga mampu meningkatkan prestasinya.

Hasil pengujian menggunakan *confussion matrix* dengan 400 data *training* dan 20 data *testing* menghasilkan nilai akurasi sebesar 95% dengan nilai *recall* sebesar 98%.

“Implementasi Algoritma Naive Bayes Dalam Penentuan Pemberian Kredit” [2] Pengajuan aplikasi kredit oleh calon nasabah sekarang sangatlah mudah, hal ini dikarenakan pengajuan kredit bisa dilakukan semua orang sepanjang memenuhi syarat tertentu. Pemberian kredit kepada nasabah adalah kegiatan rutin yang mempunyai risiko tinggi, hal ini bisa menyebabkan kerugian pada perusahaan dan mengakibatkan kredit macet. Persaingan perusahaan penyedia kredit menjadi sangat pesat dan prediksi konsumen kredit adalah hal yang sangat penting. Analisis terhadap data kredit diperlukan dengan tujuan untuk meminimalisasi risiko nasabah yang terlambat membayar kredit, kegiatan ini sangatlah penting karena salah satu penyebab terjadinya kredit macet bisa disebabkan oleh kurang cermatnya perusahaan dalam pemberian kredit.

Masalah ini sebenarnya dapat diatasi dengan cara mengidentifikasi dan memprediksi nasabah dengan baik sebelum memberikan pinjaman dengan cara memperahitkan data historis pinjaman. Teknik prediksi dalam pengambilan keputusan telah banyak digunakan oleh perusahaan-perusahaan besar. Penelitian ini menerapkan algoritma naive bayes untuk memprediksi dan mengklasifikasi nasabah mana saja yang bermasalah dan tidak bermasalah, dan diharapkan mampu meningkatkan akurasi dalam menganalisa kelayakan kredit.

Dari hasil penelitian terbukti bahwa model Naive Bayes mempunyai akurasi yang baik, hal ini tampak dari hasil evaluasi penelitian bahwa model Naive Bayes mampu menganalisa pelanggan yang baik dan pelanggan yang buruk baik menggunakan data *Agiing leasing ACC* maupun menggunakan data *credit approval negara Australia dan Japan* dari UCI data set dengan tingkat akurasi yang baik. Banyaknya record dan atribut pada sebuah data set mempengaruhi tingkat akurasi dari model Naive Bayes.

“Penerapan Algoritma Naive Bayes Untuk Penentuan Calon Penerimaan Beasiswa Pada SMK PASIM PLUS SUKABUMI” [3] Setiap lembaga khususnya sekolah banyak sekali memberikan beasiswa kepada siswanya, baik yang mampu maupun tidak mencapai hasil yang baik. Beasiswa tersebut dimaksudkan untuk membantu meringankan beban biaya bagi mahasiswa yang mendapatkannya. Dengan banyak siswa yang mendaftar untuk menerima beasiswa dan kriteria penilaian yang banyak pula, jadi tidak

semua mahasiswa yang mengajukan beasiswa dapat dikabulkan.

SMK Pasim Plus Sukabumi belum memiliki sistem yang dapat membantu dalam penentuan penerima beasiswa secara lebih efektif dan efisien. Metode algoritma Bayes diharapkan dapat membantu dalam penentuan calon penerima. Algoritma Bayes adalah salah satu metode data mining berisi sepuluh klasifikasi data mining yang paling populer di kalangan algoritma lain. Metode Naïve Bayes juga menilai potensi baik dalam mengklasifikasikan dokumen dari pada metode klasifikasi lainnya dalam hal akurasi dan efisiensi komputasi.

2.2. Landasan Teori

2.2.1. Data Mining

Data Mining didefinisikan sebagai proses penggalian yang valid, dipahami sebelumnya dan tidak diketahui, informasi dari *database* yang besar dalam rangka meningkatkan dan mengoptimalkan keputusan bisnis. Secara garis besar *Data mining* dapat dikelompokkan menjadi 2 kategori utama:

1. *Descriptive mining*, yaitu proses untuk menemukan karakteristik penting dari data dalam suatu basis data. Teknik Data mining yang termasuk dalam *Descriptive mining* adalah *clustering*, *association*, dan *sequential mining*.
2. *Predictive*, yaitu proses untuk menemukan pola dari data dengan menggunakan beberapa variable lain di masa depan. Salah satu teknik yang terdapat dalam *predictive mining* adalah klasifikasi.

Data mining adalah kegiatan menemukan pola yang menarik dari data dalam jumlah besar, data dapat disimpan dalam *database*, *data warehouse*, atau penyimpanan informasi lainnya. *Data mining* berkaitan dengan bidang ilmu – ilmu lain, seperti *database system*, *data warehousing*, *statistik*, *machine learning*, *information retrieval*, dan komputasi tingkat tinggi.

Selain itu, *Data mining* didukung oleh ilmu lain seperti *neural network*, pengenalan pola, *spatial data analysis*, *image database*, *signal processing*. *Data mining* didefinisikan sebagai proses menemukan pola-pola dalam data. Proses ini otomatis atau seringnya semi otomatis. Pola yang ditemukan harus penuh arti dan pola tersebut memberikan keuntungan, biasanya keuntungan secara ekonomi. Data yang dibutuhkan dalam jumlah besar. *Data mining* dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan, yaitu:

1. Deskripsi
Deskripsi adalah menggambarkan pola dan kecenderungan yang terdapat dalam data secara sederhana. Deskripsi dari pola dan kecenderungan sering memberikan kemungkinan penjelasan untuk suatu pola atau kecenderungan.
2. Klasifikasi
Klasifikasi suatu teknik dengan melihat pada kelakuan dan atribut dari kelompok yang telah didefinisikan. Teknik ini dapat memberikan klasifikasi pada data baru dengan memanipulasi data yang ada yang telah

diklasifikasi dan dengan menggunakan hasilnya untuk memberikan sejumlah aturan. Klasifikasi menggunakan *supervised learning*.

3. Estimasi
Estimasi hamper sama dengan klasifikasi, perbedaannya adalah variabel target estimasi lebih kearah numerik dari pada kearah kategori. Model dibangun dengan menggunakan *record* lengkap yang menyediakan nilai dari variabel target sebagai nilai prediksi.
4. Prediksi
Prediksi memiliki kesamaan dengan klasifikasi dan estimasi, dalam prediksi nilai dari hasil prediksi akan ada dimasa mendatang. Beberapa teknik yang digunakan dalam klasifikasi dan estimasi dapat juga digunakan (untuk keadaan yang tepat) untuk prediksi.
5. Klastering
Klastering merupakan pengelompokan *record*, pengamatan, atau memperhatikan dan membentuk kelas objek - objek yang memiliki kemiripan satu dengan yang lainnya dan memiliki ketidak miripan dengan *record-record* dalam kluster lain. Klastering menggunakan *unsupervised learning*.
6. Asosiasi
Tugas asosiasi atau sering disebut juga sebagai "market basket analysis" dalam Data mining adalah menemukan relasi atau korelasi diantara himpunan item-item dan menemukan atribut yang muncul dalam satu waktu. Asosiasi menggunakan *unsupervised learning*. Penting tidaknya suatu aturan asosiatif dapat diketahui dengan dua parameter, *support* dan *confidence* [4].

2.2.2. Proses Tahapan Data Mining

Data Mining adalah "proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terakit dari berbagai *database* besar". *Data mining* juga merupakan bagian dari proses *Knowledge Discovery in Database* (KDD).

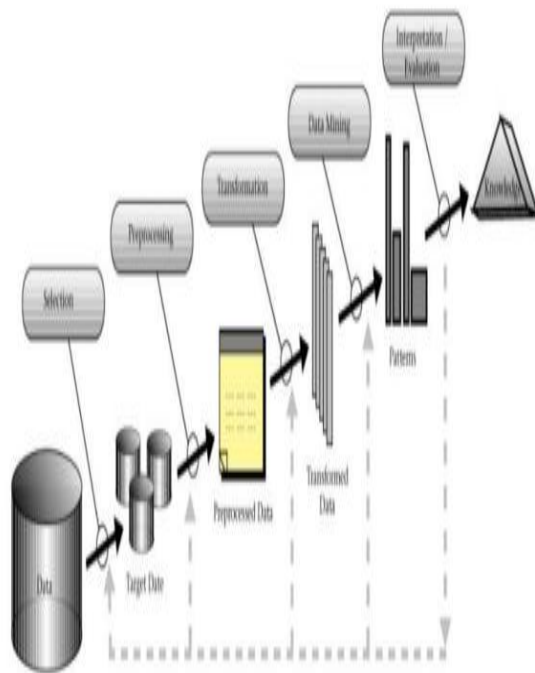
Proses KDD terdiri dari langkah-langkah berikut :

1. *Data cleaning*, menghilangkan *noise* dan data yang tidak konsisten.
2. *Data integration*, mengintegrasikan beberapa sumber data yang dapat digabungkan.
3. *Data selection*, menyeleksi data yang relevan dengan tugas analisis akan diambil dari *database*.
4. *Data transformation*, proses dimana data ditransformasikan atau dikonsolidasi dalam format yang sesuai untuk diproses dalam *data mining*.
5. *Data mining*, sebuah proses esensial dimana metode diaplikasikan untuk mengekstrak pola data.
6. *Pattern evaluation*, proses untuk mengidentifikasi pola-pola yang menarik

untuk direpresentasikan kedalam *knowledge based*.

7. *Knowledge Presentation*, proses visualisasi dan teknik representasi pengetahuan yang digunakan untuk menyajikan pengetahuan yang berguna kepada pengguna.

Proses KDD dapat digambarkan sebagai berikut :



Gambar 1. Proses *Knowledge Discovery in Database*

KDD mengacu pada proses keseluruhan yang dilakukan untuk menemukan pengetahuan yang tersimpan dari data. Proses yang terjadi dalam KDD adalah sebagai berikut :

- a. Seleksi

Pada tahap ini dilakukan pemilihan data yang berguna dalam proses *Data Mining* untuk membantu mencari solusi masalah. Karena tidak semua data yang digali akan digunakan pada tahap ini.
- b. Pra-pengolahan

Pada tahap ini dilakukan proses pra-pengolahan data, termasuk pembersihan data untuk menyingkirkan anomali dari data yang rusak, hilang, maupun mengalami kesalahan.
- c. Transformasi

Pada tahap ini dilakukan pengubahan tipe data agar sesuai dengan data yang digunakan untuk proses pengolahan.
- d. *Data Mining*

Pada tahap ini diterapkan pemilihan teknik *data mining* yang sesuai dengan tujuan dan algoritma *Data Mining* untuk pencarian pola yang menarik.
- e. Interpretasi/evaluasi

Pada tahap ini dilakukan visualisasi dan interpretasi terhadap pola yang ditemukan untuk dijadikan pengetahuan serta menghapus

pola berlebihan dari yang telah dihasilkan. Pengetahuan membantu pengguna untuk memanfaatkan pengetahuan yang diperoleh untuk mengambil keputusan yang terbaik [5].

2.2.3 Klasifikasi

Klasifikasi adalah salah satu pembelajaran yang paling umum di *data mining*. Klasifikasi didefinisikan sebagai bentuk analisis data untuk mengekstrak model yang akan digunakan untuk memprediksi label kelas. Kelas dalam klasifikasi merupakan atribut dalam satu set data yang paling unik yang merupakan variable bebas dalam statistik. Klasifikasi data terdiri dari dua proses yaitu tahap pembelajaran dan tahap pengklasifikasian. Tahap pembelajaran merupakan tahapan dalam pembentukan model klasifikasi, sedangkan tahap pengklasifikasian merupakan tahapan penggunaan model klasifikasi untuk memprediksi label kelas dari suatu data [6]

Klasifikasi (*Classification*) merupakan fungsi data mining yang memberikan item dalam koleksi untuk kategori target atau kelas. Tujuannya adalah memprediksi kelas target secara akurat untuk setiap kasus dalam data. contoh, model klasifikasi yang memprediksi resiko kredit dapat dikembangkan berdasarkan data yang diamati pada pemohon pinjaman selama beberapa periode waktu tertentu. Untuk mendapatkan data tersebut, mungkin dapat memulainya dengan melacak riwayat pekerjaan, kepemilikan rumah atau sewa, tahun tinggal, jumlah dan jenis investasi, dan seterusnya. Peringkat kredit akan menjadi target, atribut lainnya akan menjadi *prediktor*, dan data untuk setiap pelanggan akan merupakan suatu kasus.

2.2.4 Algoritma Naive Bayes

Algoritma Naive Bayes merupakan salah satu algoritma yang terdapat pada teknik klasifikasi yang menggunakan metode probabilitas yang sederhana berdasarkan pada *teorembayes* dengan asumsi tidak ketergantungan (*independent*) yang tinggi. Beberapa studi mengenai algoritma klasifikasi menunjukkan bahwa *naive bayes* memiliki performa yang sebanding dengan *decision tree* dan *neural network classifier* tertentu. Selain itu, metode ini juga menunjukkan akurasi dan kecepatan yang tinggi ketika digunakan dalam basis data yang berukuran besar.

Naive Bayes merupakan sebuah pengklasifikasian probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris *Thomas Bayes* yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya, sehingga dikenal sebagai *Teorema Bayes*. Teorema tersebut dikombinasikan dengan *naive* dimana diasumsikan kondisi antar atribut saling bebas. Klasifikasi *naive bayes* diasumsikan bahwa ada atau tidak ciri tertentu dari sebuah kelas tidak ada hubungannya dengan ciri dari kelas lainnya.

Pengertian lain dari *naive bayes* yaitu sebuah klasifikasi probabilistik sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan, algoritma menggunakan *teorema bayes* dan mengasumsikan semua atribut *independent* yang

diberikan oleh nilai pada variabel kelas. Keuntungan penggunaan metode *naive bayes* adalah bahwa metode ini hanya membutuhkan jumlah data pelatihan (*Data Training*) yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian, *naive bayes* sering bekerja jauh lebih baik dalam kebanyakan situasi dunia nyata yang kompleks dari pada yang diharapkan.

3. Metode Penelitian

3.2 Metode Pengumpulan Data

Metode pengumpulan data adalah cara atau teknik yang dapat digunakan oleh peneliti untuk mengumpulkan data. Untuk memperoleh data yang akurat dalam laporan penelitian ini, maka penulisan ini menggunakan beberapa metode yang dilakukan yaitu sebagai berikut:

3.2.1 Jenis Data

Adapun jenis data yang digunakan dalam penelitian ini adalah sebagai berikut:

1. Data Kualitatif

Data yang disajikan dalam bentuk kata verbal bukan dalam bentuk angka. Yang termasuk data kualitatif dalam penelitian ini yaitu gambaran umum obyek penelitian, meliputi: Penjelasan singkat struktur organisasi IPSI dan data calon atlet.

2. Data Kuantitatif

Data yang dapat diukur atau dihitung secara langsung, yang berupa informasi atau penjelasan yang dinyatakan dengan bilangan atau berbentuk angka. Dalam hal ini data kuantitatif yang diperlukan adalah data yang digunakan sebagai pengujian berupa angka meliputi: nilai IMT, massa otot, dan massa lemak dalam bentuk angka sebagai acuan dalam penentuan parameter pada penelitian.

3.2.2 Sumber Data

Data yang digunakan dalam penelitian ini adalah data training dan data testing. Data tersebut merupakan dari data calon atlet di organisasi pencak silat kabupaten bekasi.

Di bawah ini merupakan sample data konsumen sebelum dilakukan pengolahan data.

No	Nama	Tempat, Tanggal lahir	JK	Usia	Berat Badan	Tinggi Badan	Nilai IMT	Massa Otot	Massa Lemak	Ket
1	Ahmad Kelvin Saputra	Bekasi, 22 Mei 2008	L	13	40	137	21,311737	19	23,16	Lolos
2	Aprilia Amanda Putri S	Bekasi, 13 April 2009	P	12	40	139	20,702862	16,11	22,20	Lolos
3	Bilhaq Yalya Syuhada	Bekasi, 12 April 2008	L	13	60	148	27,392257	21	30,46	Gagal
4	Fitri Ayu	Bekasi, 12 Juli 2008	P	13	54	144	26,041667	18,53	28,84	Gagal
5	Halifat Zahra	Bekasi, 19 September 2008	P	13	41	137	21,844531	16,65	23,80	Lolos
6	Jejen Juhaeni	Bekasi, 29 Desember 2008	L	13	50	148	22,826881	19,08	24,98	Lolos
7	Leli Damayanti	Bekasi, 18 Desember 2008	P	13	42	143	20,538902	15,97	22,24	Lolos
8	Muhammad Alip	Bekasi, 10 November 2008	L	13	47	143	22,984009	19,46	25,17	Lolos
9	Mahmud Rais	Bekasi, 31 Oktober 2008	L	13	46	147	21,287427	18,38	23,13	Lolos
10	Mariyah	Bekasi, 27 Juli 2008	P	13	39	140	19,897959	15,63	21,47	Lolos
11	Muhammad Pabri	Bekasi, 06 November 2008	L	13	48	143	23,47303	19,68	25,76	Lolos
12	Nasasya Ailka Ramadhani	Bekasi, 04 Agustus 2008	P	13	52	147	24,064047	17,69	26,47	Lolos
13	Nur Aisyah Amanda	Bekasi, 28 April 2008	P	13	40	145	19,02497	15,14	20,42	Lolos
14	Reihan Nur Aqli	Bekasi, 23 Januari 2008	L	13	39	144	18,80787	17,21	20,16	Lolos
15	Reviyana Nur Fadhillah	Bekasi, 28 Maret 2008	P	13	49	147	22,675737	17,05	24,80	Lolos
16	Riki Fadi	Bekasi, 30 April 2008	L	13	50	143	24,451073	20,12	26,93	Lolos
17	Ricky	Bekasi, 10 Juni 2009	L	12	49	149	22,071078	18,7	23,85	Lolos
18	Romi Azhari	Bekasi, 22 April 2008	L	13	45	148	20,544193	17,93	22,24	Lolos
19	Sakti	Bekasi, 23 November 2009	L	12	45	130	26,627219	21,87	29,31	Gagal
20	Siti Aisyah	Bekasi, 15 November 2008	P	13	49	138	25,729889	18,4	28,47	Gagal

Gambar 2. Sample Data Calon Atlet

3.3 Data yang digunakan

Pada penelitian ini data yang digunakan adalah data calon atlet cabang olahraga pencak silat kabuapten

bekasi masih berupa raw data, yang berisi atribut nama, tempat tanggal lahir, jenis kelamin, usia, berat badan (Kg), tinggi badan (Cm), nilai IMT, massa otot, massa lemak, dan keterangan. Kemudian data yang akan digali yaitu nilai dari setiap kriteria tersebut menjadi patokan untuk menilai kelayakan calon atlet.

Dengan mengamati persoalan diatas *data mining* dengan metode *naive bayes* sangat tepat digunakan untuk menghasilkan *Knowledge* kelompok prioritas diantara ratusan data calon atlet pencak silat kasus IPSI. Adapun data yang akan dijadikan sebagai data training yaitu sebanyak 700 *record* data, data training digunakan untuk melakukan proses pelatihan algoritma terhadap *dataset* untuk menghasilkan suatu klasifikasi. Sedangkan data testing digunakan untuk melakukan pengujian terhadap proses klasifikasi yang digunakan algoritma terhadap *dataset*. Adapun atribut yang digunakan pada penelitian ini.

3.3.1 Pengolahan Data Awal

Pada tahap ini menjelaskan tentang tahap awal data mining. Data yang telah didapatkan akan diolah ke format yang dibutuhkan, pengelompokkan dan penentuan attribut, variabel. Dalam melakukan pengolahan data awal, akan di training atau di hitung dan dilakukan beberapa tahapan agar didapatkan data yang bisa digunakan untuk tahap selanjutnya.

3.3.2 Data Selection

Tahap data *selection* atau pemilihan data merupakan tahap pemilihan variabel dari data yang akan dianalisis, karena tidak semua data yang terdapat dalam data mentah akan digunakan. Sehingga didapat beberapa variabel yang akan dijadikan sebagai atribut.

Table 3.1 Pemilihan Variabel

Variabel	Keterangan	Indikator	Detail Penggunaan
X1	Nama	V	Digunakan Atribut
X2	Tempat, tanggal lahir	X	-
X3	Jenis Kelamin	V	Digunakan Atribut
X4	Usia	V	Digunakan Atribut
X5	Berat Badan (Kg)	X	-
X6	Tinggi Badan (Cm)	X	-
X7	Nilai IMT	V	Digunakan Atribut
X8	Massa Otot	V	Digunakan Atribut
X9	Massa Lemak	V	Digunakan Atribut
X10	Keterangan	V	Digunakan Label

Tabel diatas menerangkan variabel yang akan digunakan dan tidak digunakan dalam penelitian ini. Indikator “V” adalah menandakan bahwa variabel tersebut digunakan, sedangkan untuk indikator “X”

adalah menandakan bahwa variabel tersebut dieliminasi atau tidak digunakan pada tahap penentuan kriteria. Pengeliminasian beberapa variabel tersebut berdasarkan nilai model yang relatif sama dan tidak mempengaruhi hasil dari proses penilaian.

3.3.3 Pre-Processing (data cleaning)

Tahap *pre-processing* merupakan tahap dilakukannya pembersihan terhadap data *missing value* yaitu data yang tidak konsisten atau kosong, dan juga dilakukannya pemilihan variabel terhadap data yang ingin digunakan pada proses *data mining*.

3.3.4 Data Integration dan Transformation

Untuk meningkatkan akurasi dan efisiensi algoritma. Data dalam penelitian ini dikategorikan atau diklasifikasi. Setelah mendapatkan data selanjutnya adalah mengklasifikasi atribut-atribut karena berpengaruh terhadap hasil prediksi. Klasifikasi data ini dilakukan dengan bantuan *tool Ms.Excel*.

Pengujian dan Validasi Hasil

Pengujian metode dilakukan untuk mengetahui hasil perhitungan yang dianalisa dan untuk mengetahui apakah fungsi bekerja dengan baik atau tidak. Setelah data dihitung secara manual, kemudian data diuji menggunakan *tools Rapid Miner* untuk memastikan apakah hasil perhitungan berjalan baik dalam menentukan calon atlet, dari hasil yang diperoleh *Rapid Miner*.

Hasil pengujian akan di validasi dan kemudian di evaluasi. Tahapan evaluasi yang dilakukan dalam penelitian ini adalah untuk memberikan penilaian dari hasil penggunaan algoritma *naive bayes* untuk mengukur hasil prediksi calon penerimaan atlet pencak silat. Nilai akurasi yang dihasilkan. untuk mengetahui tingkat *accuracy, precision, recall*.

Table 1. Perangkat Lunak (*Software*)

Software	Versi	Keterangan
Microsoft Windows	10	Sebagai sistem operasi
Microsoft Office Word	2016	Sebagai media penulisan laporan
Microsoft Office Excel	2016	Sebagai media pengolahan dataset
Rapid Miner Studio	9.1	Sebagai <i>tools</i> yang digunakan untuk mengolah dataset dan pengujian terhadap algoritma yang digunakan

Table 2. Perangkat Keras (*Hardware*)

Spesifikasi Hardware	Keterangan
Processor	Intel Core i5
Random Acces Memory (RAM)	8.00 GB
System Type	64-bit Operating System

4. Pembahasan

4.1 Dataset

Pada penelitian ini jumlah data *dataset* yang digunakan sebanyak 700 data. Data calon atlet pencak silat IPSI kabupaten bekasi. Jumlah data yang diperoleh dari proses tersebut sebanyak 700 data, yang terdiri dari atribut “Lolos” sebanyak 504 data, dan “Gagal” sebanyak 196 data.

Nama	JK	Usia	Nilai IMT	Massa Otot	Massa Lemak (%)	Ket
Ahmad Kelvin Saputra	L	13	Normal	Kurus	Kelebihan Berat	Lolos
Aprilia Amanda Putri S	P	12	Normal	Kurus	Berat Sehat	Lolos
Bilhaq Yahya Syuhada	L	13	Gemuk	Alami	Gemuk	Gagal
Fitri Ayu	P	13	Gemuk	Alami	Kelebihan Berat	Gagal
Halifatu Zahra	P	13	Normal	Normal	Berat Sehat	Lolos
Jejen Juhaeni	L	13	Normal	Kurus	Kelebihan Berat	Lolos
Leli Damayanti	P	13	Normal	Kurus	Berat Sehat	Lolos
Muhammad Alip	L	13	Normal	Kurus	Kelebihan Berat	Lolos
Mahmud Rais	L	13	Normal	Kurus	Kelebihan Berat	Lolos
Mariyah	P	13	Normal	Kurus	Berat Sehat	Lolos
Muhammad Pahri	L	13	Normal	Kurus	Gemuk	Lolos
Natasya Alika Ramadhani	P	13	Normal	Normal	Kelebihan Berat	Lolos
Nur Aisyah Amanda	P	13	Normal	Kurus	Malnutrisi	Lolos
Reihan Nur Aqli	L	13	Normal	Kurus	Kelebihan Berat	Lolos

Gambar 3. Dataset

4.1.1 Perhitungan Naive Bayes

Tabel 3. Naive Bayes

Naive Bayes				
Tahap I P(Ci)				
Langkah	Jumlah Kasus	Lolos	Gagal	Jumlah
Total	700	504		0.72 %
			196	0.28 %
Hitung P(X Ci) untuk setiap kelas				
Usia				
12		111		0.15857 %
			60	0.08571 %
13		393		0.56142 %
			136	0.19428 %
Nilai IMT				
Kurus		22		0.03142 %
			-	
Norma I		481		0.68714 %
			142	0.20285 %

	Gemuk		1		0.014 2 %
				182	0.2 6 %
Massa Otot					
	Kurus		337		0.4 814 2 %
				13	0.0 185 7 %
	Normal		140		0.2 %
				40	0.0 571 4 %
	Alami		23		0.0 328 5 %
				91	0.1 3 %
	Besar		4		0.0 057 1 %
				52	0.0 742 8 %
Massa Lemak					
	Malnutrisi		16		0.0 228 5 %
				-	
	Berat Sehat		144		0.0 257 1 %
				53	0.0 757 1 %
	Kelebihan Berat		250		0.3 571 4 %
				94	0.1 342 8 %
	Gemuk		94		0.1 342 8 %
				49	0.0 7 %

Sumber : Penulis, 2022

Tahap 3 P(X|Ci)*P(Ci)

$P(X|Klasifikasi = \text{“Lolos”} = P(\text{Usia} = \text{“13”}, \text{Nilai IMT} = \text{“Normal”}, \text{Massa Otot} = \text{“Normal”}, \text{Massa Lemak} = \text{“Berat Sehat”}) | P(Klasifikasi = \text{“Lolos”})$
0.56142 x 0.68714 x 0.2 x 0.02571 = 0.0019836

$P(X|Klasifikasi = \text{“Gagal”} = P(\text{Usia} = \text{“13”}, \text{Nilai IMT} = \text{“Normal”}, \text{Massa Otot} = \text{“Normal”}, \text{Massa Lemak} = \text{“Berat Sehat”}) | P(Klasifikasi = \text{“Gagal”})$
0.19428 x 0.20285 x 0.05714 x 0.07571 = 0.0001704

Hasil pendefinisian *Probabilitas prior* berdasarkan persamaan menghasilkan nilai untuk kelas **Lolos** sebesar **0.0019836** dan untuk kelas **Gagal** sebesar **0.0001704**.

4.2 Implementasi Klasifikasi Naive Bayes

Setelah melakukan perhitungan manual selanjutnya pembuktian dengan menggunakan *Rapid Miner*. Pengujian ini untuk mengetahui nilai akurasi dan prediksi data. Pengujian dilakukan dengan *algoritma Naive Bayes* dan data yang digunakan sebanyak 500 *data training* dan 200 *data testing*. Berikut tahap pengujiannya :

Penerapan Metode *Naive Bayes* dengan *Rapid Miner*

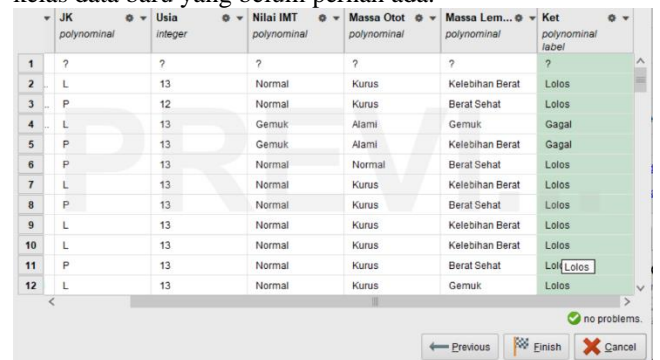
Rapid Miner Studio 9.1 adalah salah satu *tool* untuk melakukan prediksi dan analisa *data mining*. Berikut adalah tahapan dalam melakukan *data mining* di *Rapid Miner 9.1*.

Proses Import Data

Pada tahapan ini data hasil data calon atlet pencak silat yang sudah diklasifikasi, *integration* dan *selection* di *import* kedalam *tool Rapid Miner 9.1*. yaitu terdiri dari *data training* dan *data testing*.

1. Import data training

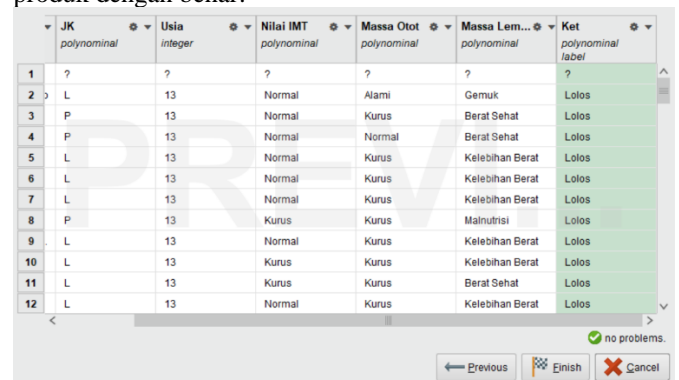
Data training ini digunakan untuk latihan membentuk sebuah model *classifier* untuk prediksi kelas data baru yang belum pernah ada.



Gambar 4. Import data training pada Rapid Miner

2. Import data testing

Data testing ini digunakan untuk mengukur sejauh mana *prediction* berhasil melakukan prediksi kualitas produk dengan benar.



Gambar 5. Import data testing pada Rapid Miner

5.1 Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan mengenai prediksi penerimaan Calon atlet pusat pendidikan dan latihan cabang olahraga pencak

silat dengan metode *naive bayes* maka dapat diambil beberapa kesimpulan sebagai berikut :

1. Penerapan metode *naive bayes* memanfaatkan *data training* untuk menghasilkan *probabilitas* setiap kriteria untuk *class* yang berbeda, sehingga nilai-nilai *probabilitas* dari kriteria tersebut dapat dioptimalkan untuk menentukan prediksi penerimaan calon atlet pusat pendidikan dan latihan remaja cabang olahraga pencak silat “ Lolos “ dan “Gagal “ dengan cepat dan efisien berdasarkan klasifikasi yang dilakukan oleh metode *naive bayes*.
2. Dari hasil pengujian yang telah dilakukan maka hasil penelitian dengan metode *naive bayes* pada *Rapid Miner* maka didapatkan sebuah hasil bahwa nilai akurasi adalah 89.24 %. Sehingga hasil ini adalah termasuk akurasi yang sangat baik. Dengan begitu metode ini bisa diterapkan dalam memprediksi penerimaan calon atlet pencak silat.
3. Hasil perbandingan dari dua metode yang dilakukan oleh penulis menghasilkan kesimpulan data yang digunakan lebih cocok diproses menggunakan algoritma *naive bayes*. Karena data yang ada apabila diproses oleh algoritma C4.5 akan memproses setiap kriteria yang ada sehingga tidak efisien dalam menentukan calon atlet yang “Lolos” dan “Gagal”. Sedangkan algoritma *naive bayes* memproses seluruh kriteria yang telah di klasifikasikan sehingga lebih efisien untuk menentukan “Lolos” atau “Gagal” calon atlet pencak silat.

Daftar Pustaka

- [1] J. Tedy Setiadi, “Penerapan Klasifikasi Bayes Untuk Memprediksi Jenis Latihan Siswa Pencak Silat (Studi Kasus Pencak Silat PSHT),” *TEKNIKA*, vol. 7, 2018.
- [2] A. W. Muhammad Husni Rifqo, “Implementasi Algoritma Naive Bayes Dalam Penentuan Pemberian Kredit,” *Pseudocode*, vol. IV, no. 2, 2017.
- [3] S. A. Rizal Amegia Saputra, “Penerapan Algoritma Naive Bayes Untuk Penentuan Calon Penerimaan Beasiswa Pada SMK Pasim Plus Sukabumi,” *SWABUMI*, no. 2, p. IV, 2016.
- [4] d. T. P. Bijak Jati Kusuma, “Rancang Bangun Aplikasi Mobile Perhitungan Indeks Massa Tubuh dan Berat Badan Ideal,” pp. 157-158, 4 November 2011.
- [5] D. L. Fithri, “Model Data Mining Dalam Penentuan Kelayakan Pemilihan Tempat Tinggal Menggunakan Metode Naive Bayes,” *jurnal SIMETRIS*, pp. 727-728, 2016.
- [6] D. E. P. Eka Praja Wiyata Mandala, “Prediksi Jumlah Pemberian Kredit Kepada Nasabah di Bank Perkreditan Rakyat Dengan Algoritma C 4.5,” *KomTekInfo*, pp. 71-72, 2018.
- [7] D. I. S. Dewi Sartika, “Perbandingan Algoritma Klasifikasi Naive Bayes, Nearest Neighbour, dan Decision Tree pada Studi Kasus Pengambilan Keputusan Pemilihan Pola Pakaian,” *JatISI*, p. 153, 2017.
- [8] A. B. A. Supangat, “SISTEM INFORMASI SEKOLAH MENGGUNAKAN ALGORITMA C4.5 UNTUK MENGIDENTIFIKASI FAKTOR FAKTOR PENURUNAN MUTU PEMBELAJARAN SISWA SMK,” pp. 5-6, 7 JULI 2020.
- [9] A. S. E. S. Siska Haryati, “Implementasi Data Mining Untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma C 4.5,” *Jurnal Media Infotama*, p. 133, 2015.
- [10] B. Syahid, “Pencak Silat,” *GuruPendidikan.com*, 2021.
- [11] “Kalkulator FFMI (Fat Free Mass Index),” *kalkulator.id*, 31 juli 2021.