



PERBANDINGAN TIGA ALGORITMA CLASSIFIER UNTUK PENENTUAN PENERIMAAN PESERTA DIDIK BARU PADA SEKOLAH MENENGAH ATAS

Yulhendri^a, Kundang Karsono Juman^b

^a Fakultas Ilmu Komputer / Sistem Informasi, yulhendri@esaunggul.ac.id, Universitas Esa Unggul

^b Fakultas Ilmu Komputer / Teknik Informatika, kundang.karsono@esaunggul.ac.id, Universitas Esa Unggul

Corresponding Author : yulhendri@esaunggul.ac.id

ABSTRAK

New Student Admissions (PPDB) have been carried out since the 2012/2013 school year by schools that have been verified by the Ministry of Education. From the 2012/2013 school year to the 2018/2019 school year, 1279 schools have been registered and verified. The activity of accepting new students (PPDB) for SMA is not an independent one, but is not separated from various aspects. Education staff, participants, schools, school quotas, and participant scores. And one of the most important processes of new student admissions is the prediction of the selection of high school schools. The problem faced by prospective students is prediction for school selection. The predictions made so far have only focused on the passing grade of each school. However, the passing grade that is focused on only revolves around the previous year. Therefore, we need a system to predict participants in choosing schools from new student admissions activities by implementing several machine learning algorithms.

Keywords: Machine Learning, Prediction, PPDB High School.

Abstrak (Times New Roman 10, Bold, spasi 1, spacing before 12 pt, after 2 pt)

Penerimaan Peserta Didik Baru Baru (PPDB) sudah dilakukan sejak tahun ajaran 2012/2013 oleh sekolah yang sudah diverifikasi oleh Departemen Pendidikan. Dari tahun ajaran 2012/2013 sampai tahun ajaran 2018/2019 sudah 1279 sekolah yang sudah terdaftar dan sudah diverifikasi. Kegiatan Penirimaan peserta didik baru (PPDB) SMA bukan yang berdiri sendiri, namun tidak dipisahkan dari berbagai aspek. Tenaga penddika, peserta, sekolah, kuota sekolah, dan nilai peserta. Dan salah satu proses yang terpenting dari kegiatan penerimaan peserta didik baru adalah prediksi pemilihan sekolah SMA. Masalah yang dihadapi oleh para calon siswa adalah prediksi untuk pemilihan sekolah. Prediksi yang dilakukan sampai sekarang hanya terfokus pada passing grade setiap sekolah. Akan tetapi, passing grade yang di fokuskan hanya berkisar di tahun sebelum nya. Oleh karena itu, di butuhkan sebuah sistem untuk memprediksikan peserta dalam memilih sekolah dari kegiatan penerimaan peserta didik baru dengan mengimplementasi dari Beberapa Algoritma Machine Learning.

Kata Kunci: *Machine Learning*, Prediksi, PPDB SMA.

1. PENDAHULUAN

Pesatnya kemajuan zaman menurut semua orang selalu berkembang dalam pertukaran informasi dari berbagai sumber. Dengan memanfaatkan kemajuan teknologi tersebut, di implementasikan segi Pelayanan Pendidikan. Penerimaan Peserta Didik baru SMA dimaksudkan untuk memberikan layanan pendidikan seluas-luasnya nya bagi masyarakat . Hal tersebut sejalan dengan wajib belajar 12 tahun. Kegiatan Penirimaan peserta didik baru (PPDB) SMA bukan yang berdiri sendiri, namun tidak dipisahkan dari berbagai aspek. Tenaga penddika, peserta, sekolah, kuota sekolah, dan nilai peserta. Dan salah satu proses yang terpenting dari kegiatan penerimaan peserta didik baru adalah prediksi pemilihan sekolah SMA. Penerimaan Peserta Didik Baru (PPDB) SMA yang dilaksanakan secara Online bertujuan untuk menjamin penerimaan peserta didik baru berjalan secara objektif, transparan, akuntabel, non-diskriminatif, dan

berkeadilan dalam rangka mendorong peningkatan akses layanan pendidikan. Selain itu PPDB juga menjadi upaya pemerintah untuk pemerataan pendidikan, sehingga tidak terdapat lagi beberapa sekolah yang sangat diminati oleh pendaftar, sedangkan di beberapa sekolah lainnya kurang peminat. Penerimaan Peserta Didik Baru Baru (PPDB) sudah dilakukan sejak tahun ajaran 2012/2013 oleh sekolah yang sudah diverifikasi oleh Departemen Pendidikan. Dari tahun ajaran 2012/2013 sampai tahun ajaran 2018/2019 sudah 1279 sekolah yang sudah terdaftar dan sudah diverifikasi. Pada SMAN 112 Jakarta Barat, SMAN 65 Jakarta Barat, SMAN 57 Jakarta Barat sudah termasuk oleh sekolah yang terdaftar. Di SMAN 112 Jakarta terdapat 627 siswa terdaftar dari tahun 2012-2018 sedangkan untuk proses kegiatan PPDB dengan 2 tahap, tahap pertama untuk domisili Jakarta barat (Umum), tahap 2 untuk luar domisili Jakarta barat (Umum). Di SMAN 65 Jakarta Barat terdapat 461 siswa terdaftar dari tahun 2012-2018, sama halnya dengan SMAN 112 Jakarta barat, SMAN 65 juga melakukan 2 tahap, tahap pertama untuk domisili Jakarta barat (Umum), tahap 2 untuk luar domisili Jakarta barat (Umum). Pada SMAN 57 terdapat 512 siswa terdapat 514 siswa yang terdaftar dari tahun 2012-2018. (dalam hasil observasi arsip data dari tahun ajaran 2012/2014).

Masalah yang dihadapi oleh para calon siswa adalah prediksi untuk pemilihan sekolah. Prediksi yang dilakukan sampai sekarang hanya terfokus pada passing grade setiap sekolah. Akan tetapi, passing grade yang di fokuskan hanya berkisar di tahun sebelumnya. Passing grade adalah nilai terendah setiap sekolah di tahun sebelumnya. Calon peserta banyak yang memfokuskan pada passing grade setiap sekolah, tidak melihat data-data siswa yang mendaftar tahun sebelumnya. Calon peserta memilih sekolah yang di inginkan sedangkan nilai akhir ujian nasional tidak mencukupi, dan kehilangan satu kesempatan untuk masuk sekolah negeri. Oleh karena itu, di butuhkan sebuah sistem untuk memprediksikan peserta dalam memilih sekolah dari kegiatan penerimaan peserta didik baru dengan mengimplementasi dari Beberapa Algoritma *Machine Learning* yaitu *Support Vector Machine (SVM)*, *Naïve Bayes*, dan *Decision Tree*.

2. TINJAUAN PUSTAKA

2.1. *Machine Learning*

Machine Learning (Pembelajaran Mesin) merupakan kemampuan komputer untuk melakukan pembelajaran tanpa harus menjelaskan (programmed) secara eksplisit kepada komputer, atau menurut Tom Mitchel suatu komputer dikatakan melakukan pembelajaran dari pengalaman E terhadap tugas T dan mengukur peningkatan kinerja P, jika kinerja Tugas T diukur oleh kinerja P, meningkatkan pengalaman E.

Supervised Learning memiliki karakteristik masalah yang diselesaikan biasanya berupa klasifikasi, dataset yang dimiliki oleh kasus yang berbentuk klasifikasi biasanya selain memiliki atribut untuk setiap instancesnya namun juga sudah memiliki kelas yang jelas, sehingga task selanjutnya dari hipotesis atau model yang ditemukan adalah melakukan klasifikasi terhadap instance yang baru dan belum memiliki label.

Unsupervised Learning biasanya memiliki kata kunci clustering atau melakukan pengklusteran terhadap sekelompok data atau sekelompok instances yang tidak memiliki label, sehingga memiliki informasi bahwa terdapat sekumpulan data yang membentuk cluster, namun kita belum tahu apa pengetahuan atau hipotesis yang membuat instances tersebut saling berkumpul (membuat kelompok) menjadi satu cluster atau lebih. Contohnya seperti K-Mean, Graph Algorithm, EM (Expectation Maximization) Algorithm.

Sedangkan *Reinforcement Learning* biasanya berupa permasalahan yang membutuhkan aktifitas eksplorasi, sehingga cukup sesuai jika digunakan untuk membangun suatu intelijen pada suatu game (terutama puzzle).

2.1.1 Langkah-Langkah Dalam Mengembangkan Aplikasi *Machine Learning* (Harrington, 2015).

- a. Proses pengumpulan data dilakukan dengan mengambil contoh dari berbagai sumber informasi, seperti di Internet dan media cetak. Data yang dikumpulkan adalah data yang disebar secara bebas ke publik.
- b. Mempersiapkan Data Masukan Pada hal ini data masukan yang disiapkan adalah data masukan yang sesuai dengan format yang dibutuhkan untuk analisis.
- c. Menganalisis Data Masukan Setelah proses pertama dan kedua dilakukan, maka hal selanjutnya yang harus dilakukan adalah menganalisis data masukan dan untuk menganalisis dapat dilakukan

dengan melihat pola data dan juga dengan memisahkan data berdasarkan dimensi masing-masing data.

- d. Mengikutsertakan Keterlibatan Manusia
- e. Melatih Algoritma Pada langkah ini pengguna “memberi makan” algoritma dengan data yang berkualitas, dan nantinya algoritma akan mengolah data tersebut menjadi informasi serta menyimpannya.
- f. Menguji Algoritma Pada langkah ini hal yang dilakukan adalah melihat seberapa baik kualitas algoritma yang telah dilatih pada tahap sebelumnya.
- g. Menggunakannya Langkah ini merupakan langkah akhir untuk algoritma yang diterapkan dalam suatu program, sehingga dapat melakukan suatu hal. Kemudian dilakukan pengecekan ulang terhadap langkah-langkah sebelumnya.

2.2 Algoritma Machine Learning

2.2.1 Support Vector Machine

(SVM) SVM adalah sebuah algoritma yang bekerja menggunakan pemetaan nonlinear untuk mengubah data pelatihan asli ke dimensi yang lebih tinggi, dalam dimensi yang baru, kemudian akan mencari linear optimal pemisah hyperplane (yaitu, “decision boundary” yang memisahkan tupel dari satu kelas dengan kelas lainnya). Dengan pemetaan nonlinear yang tepat untuk dimensi yang cukup tinggi, data dari dua kelas selalu dapat dipisahkan dengan hyperplane. SVM menemukan hyperplane ini menggunakan support vector (“essential” training tuples) dan margins (didefinisikan oleh support vectors).

2.2.2 Naïve Bayes

Naïve Bayes merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes. *Naïve Bayes* untuk setiap kelas keputusan, menghitung probabilitas dengan syarat bahwa kelas keputusan adalah benar, mengingat vektor informasi obyek. Algoritma ini mengasumsikan bahwa atribut obyek adalah independen. Probabilitas yang terlibat dalam memproduksi perkiraan akhir dihitung sebagai jumlah

2.2.3 Decesion Tree

Decesion Tree sendiri merupakan metode klasifikasi dan prediksi yang sangat kuat dan banyak diminati . Dalam decision tree ini data yang berupa fakta dirubah menjadi sebuah pohon keputusan yang berisi aturan dan tentunya dapat lebih mudah dipahami dengan bahasa alami. Model pohon keputusan banyak digunakan pada kasus data dengan output yang bernilai diskrit .

2.3 Rapid Miner

Rapid Miner merupakan software/perangkat lunak untuk pengolahan data, dengan menggunakan prinsip dan algoritma data mining, Rapidminer mengekstrak pola-pola dari data set yang besar dengan mengkombinasikan metode statistika, kecerdasan buatan dan database. Rapidminer memudahkan penggunaannya dalam melakukan perhitungan data yang sangat banyak dengan menggunakan operator-operator. Operator ini berfungsi untuk memodifikasi data. Data dihubungkan dengan node-node pada operator kemudian pengguna hanya tinggal menghubungkannya ke node hasil untuk melihat hasilnya. Hasil yang diperlihatkan RapidMiner pun dapat ditampilkan secara visual dengan grafik menjadikan RapidMiner adalah salah satu software pilihan untuk melakukan ekstraksi data dengan metode-metode data mining (Rahmat C.T.I. et al., 2017). Rapid Miner adalah aplikasi data mining yang tidak perlu dipertanyakan lagi dan berbasis sistem open-source dunia yang terkemuka dan ternama. Tersedia sebagai aplikasi yang berdiri sendiri untuk analisis data dan sebagai mesin data mining untuk integrasi ke dalam produk sendiri. Ribuan aplikasi RapidMiner di lebih dari 40 negara memberikan pengguna mereka keunggulan yang kompetitif. Solusi yang di usung antara lain :Integrasi data, Analitis ETL, Data Analisis, dan Pelaporan dalam satu suite tunggal. Powerfull tapi memiliki antarmuka pengguna grafis yang intuitif untuk desain analisis proses.Repositori untuk proses, data dan penanganan meta data Hanya solusi dengan transformasi meta data: lupakan trial and error dan memeriksa hasil yang telah di inspeksi selama desain.

Hanya solusi yang mendukung on-the-fly kesalahan dan dapat melakukan perbaikan dengan cepat lengkap dan fleksibel: ratusan loading data, transformasi data, pemodelan data, dan metode visualisasi data. Rapid

Miner menyediakan prosedur data mining dan Machine Learning termasuk: ETL (extraction, transformation, loading), data preprocessing, visualisasi, modelling dan evaluasi. Proses data mining tersusun atas operator-operator yang nestable, dideskripsikan dengan XML, dan dibuat dengan GUI. Ditulis dalam bahasa pemrograman Java. Mengintegrasikan proyek data mining Weka dan statistika R. (Setiawan, 2016)

2.4 Metode Analisis

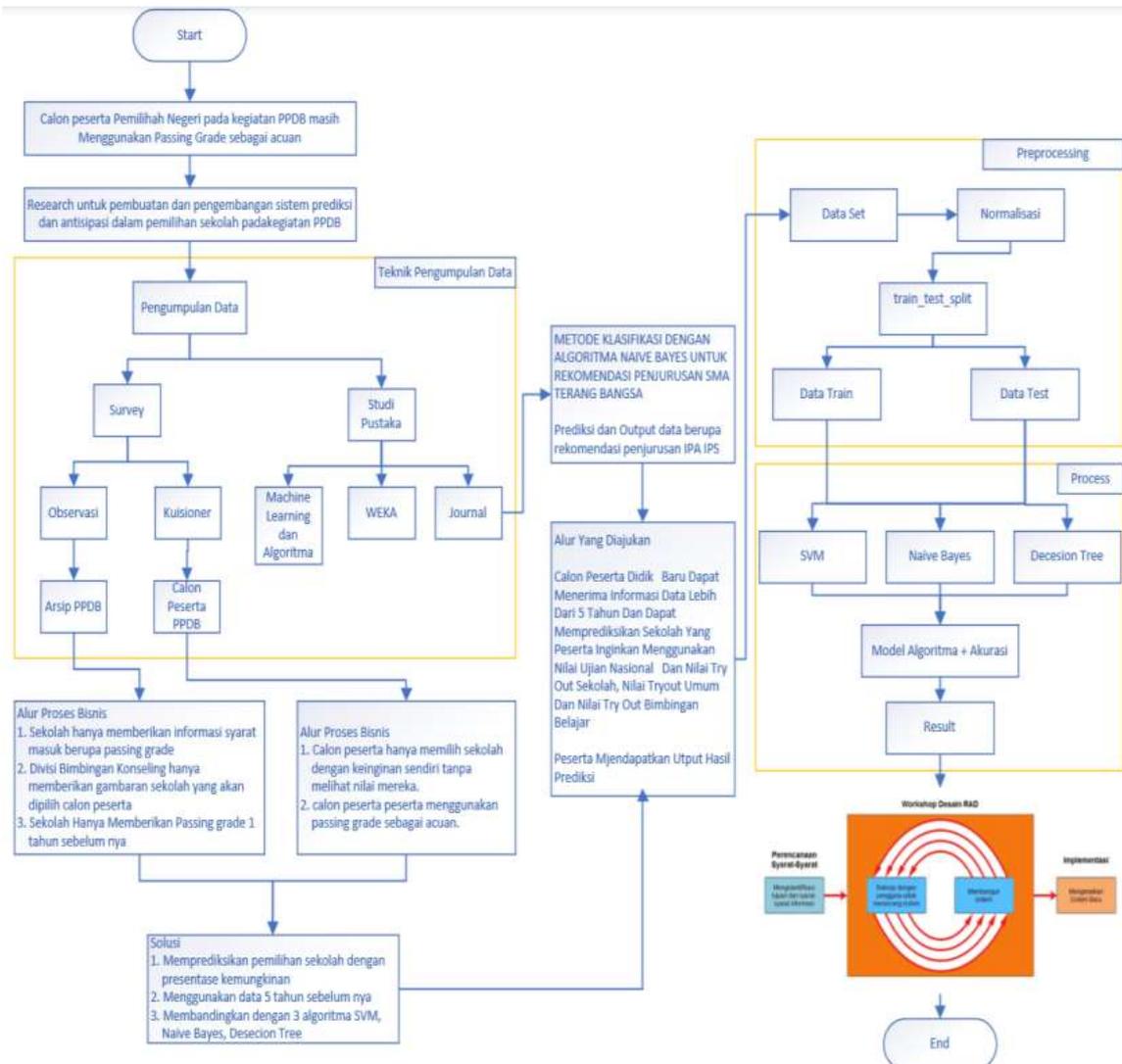
Metode analisis yang digunakan pada penelitian yang menggunakan pendekatan kualitatif tidak menggunakan alat statistik, namun dilakukan dengan menginterpretasi tabel-tabel, grafik-grafik, atau angka-angka yang ada kemudian melakukan uraian dan penafsiran. Sedangkan Analisis data secara Kuantitatif adalah metode analisis yang digunakan pada penelitian dengan pendekatan analisis kuantitatif dan menggunakan alat statistik. Dalam penelitian ini digunakan analisis kuantitatif dengan metode fishbone. Menurut (Handriani et al., 2019) "Diagram Fishbone sering juga disebut dengan istilah Diagram Ishikawa. Penyebutan diagram ini sebagai Diagram Ishikawa pada sekitar Tahun 1960-an. Diagram fishbone merupakan suatu alat visual untuk mengidentifikasi, mengeksplorasi, dan secara grafik menggambarkan secara detail semua penyebab yang berhubungan dengan suatu permasalahan. menurut dalam jurnal (Khodijah and Labibah, 2019) "Digram fishbone digunakan ketika kita ingin mengidentifikasi kemungkinan penyebab masalah dan terutama ketika sebuah team cenderung jatuh berpikir pada rutinitas." Menurut dalam jurnal (Yuniarto, 2014), "konsep dasar dari diagram fishbone adalah permasalahan mendasar diletakkan pada bagian kanan dari diagram atau pada bagian kepala dari kerangka tulang ikannya." Selain digunakan untuk mengidentifikasi masalah dan menentukan penyebabnya, diagram fishbone ini juga dapat digunakan pada proses perubahan. menyatakan diagram fishbone ini dapat diperluas menjadi diagram sebab dan akibat (cause and effect diagram). Diagram fishbone dapat digunakan untuk menganalisis permasalahan baik pada level individu, tim, maupun organisasi.

3. METODOLOGI PENELITIAN

3.1 Metodologi Penelitian

Penelitian ini dilakukan untuk mendapatkan hasil akurasi dan hasil prediksi pada kegiatan penerimaan peserta didik baru sekolah menengah atas. Prediksi disini menggunakan 3 algoritma Machine Learning diantaranya Algoritma Support Vector Machine(SVM), Algoritma Naïve Bayes, dan Algoritma Decision Tree. Dari Ketiga Algoritma tersebut di lihat mana Algoritma yang terbaik dengan tingkat akurasi tertinggi dan akan di implementasikan pada pengembangan sistem. Metodologi penelitian diawali dengan data yang digunakan, arsitektur umum penelitian, alur pengklasifikasian, dan analisis kinerja pengklasifikasian.

1. Tahapan Identifikasi Masalah Pada tahap ini penulis menganalisis masalah yang akan diselesaikan dari permasalahan yang terjadi dan dapat memecahkan masalah dan memberikan solusi dengan sistem yang diusulkan dan akan dibuat.
2. Problem Statement Fokus dari perhatian peneliti dalam menjalankan proses penelitiannya. Problem Statement adalah gambaran yang jelas tentang isu yang ingin diselesaikan atau diperbaiki melalui penelitian. Sehingga Problem Statement haruslah tersambung dengan latar belakang penelitian dan objek penelitian, atau dengan kata lain memiliki benang merah yaitu penerapan atau pembangunan data mart untuk mendukung pengambilan keputusan.
3. Teknik Pengumpulan Data Pada tahap ini ada beberapa cara yang digunakan untuk mendapatkan dan mengumpulkan data yaitu dengan survey dan studi pustaka. Ada 2 Teknik pengumpulan data yang penulis gunakan diantaranya nya Survey dan Studi Pustaka.
 - a. Survey Tahap ini adalah teknik untuk mendapatkan data yang relevan dan lengkap dengan cara observasi dan kuisisioner. Dalam tahap pertama yaitu observasi dilakukan pada arsip PPDB dengan sampel 3 sekolah (SMAN 112, SMAN 65, SMAN 57) dan menggunakan Nilai akhir, nilai Bahasa Indonesia, Nilai Matematika, Nilai Bahasa Inggris dan Nilai IPA. Tahap kedua yaitu Kuisisioner dilakukan secara pembuatan kuisisioner dan di distribusikan dengan 30 responden dengan target responden calon peserta PPDB .
 - b. Studi Pustaka Studi Pustaka: tahap ini adalah teknik untuk mendapatkan informasi dari kutipan kata yang digunakan sebagai sumber acuan untuk penelitian. Penulis Menggunakan beberapa jurnal sebagai acuan dan beberapa dokumentasi beberapa Algoritma dan Software sebagai penunjang dalam penelitian.



Gambar 1. Metodologi Penelitian

4. Proses Bisnis dan solusi Pada Tahap ini dilakukan Analisa proses bisnis dari hasil abservasi dan kuisiner. Hasil observasi dan kuisiner diperoleh kesimpulan, maka dapat memberikan solusi pada proses bisnis.

5. Alur Yang Diajukan Pada tahap ini, solusi dan journal acuan diperoleh kesimpulan untuk dikembangkan dengan data yang diperoleh dari tahap sebelum nya, maka akan diajukan alur / proses. 6. Machine Learning Arsitektur umum penelitian adalah Machine Learning untuk memprediksi menggunakan 3 algoritma Machine Learning yang berupa (SVM, Naive Bayes, Naive Bayes) dari arsip PPDB. Dalam Algoritma Machine Learning terdapat Klasifikasi dan regresi untuk membandingkan 3 algoritma dengan tingkat akurasi tertinggi yang nantinya akan di implementasikan untuk mengembangkan sistem “Machine Learning Prediksi PPDB”.

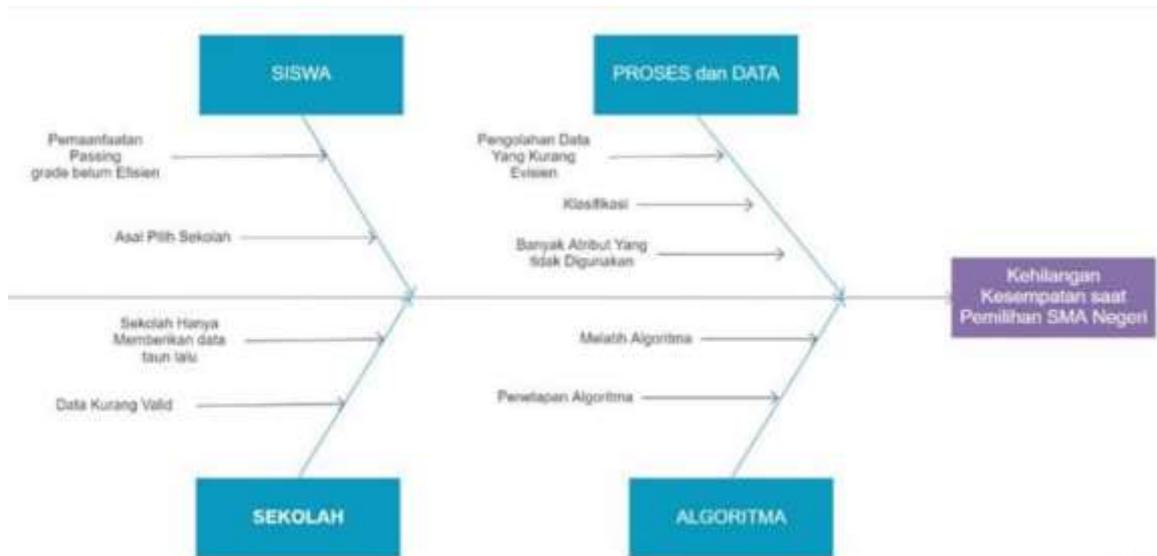
1) Preprocessing Proses awal yaitu melakukan pengumpulan data yang berkaitan dengan data yang akan diteliti, kemudian dilakukan normalisasi serta transformasi data sehingga menghasilkan data yang siap diuji. Setelah data awal / data set di normalisasi (zero-mean), selanjutnya dataset dibagi menjadi data train dan data test sehingga menghasilkan data train dan data test yang siap uji.

2) Process Proses Selanjutnya data train dan data test, di uji menggunakan 3 algoritma diantaranya Algoritma SVM, Naive Bayes, Naive Bayes, untuk membandingkan tingkat akurasi diantara tiga Algoritma Machine Learning dan menhasilkan model prediksi yang akan di implementasikan pada metode pengembangan sistem RAD yang akan menjadi system “Machine Learning Prediksi PPDB”.

PERBANDINGAN TIGA ALGORITMA CLASSIFIER UNTUK PENENTUAN PENERIMAAN PESERTA DIDIK BARU PADA SEKOLAH MENENGAH ATAS (Yulhendri)

3.2 Analisis Permasalahan

Analisis yang digunakan untuk menganalisa masalah yang ada pada Kegiatan Penerimaan Peserta Didik Baru agar dapat mengenali penyebab masalah dengan menggunakan metode Fishbone, maka dapat diperoleh beberapa penyebab masalah yang akhirnya dapat disimpulkan membantu dalam membuat rancang bangun sistem baru yang lebih baik, inilah diagram fish bone.



Gambar 2. Analisis Fish Bone

4. HASIL ANALISIS DAN PEMBAHASAN

4.1 Hasil Analisis Permasalahan

Analisis yang digunakan untuk menganalisa masalah yang ada pada Kegiatan Penerimaan Peserta Didik Baru agar dapat mengenali penyebab masalah dengan menggunakan metode Fishbone, maka dapat diperoleh beberapa penyebab masalah yang akhirnya dapat disimpulkan membantu dalam membuat rancang bangun sistem baru yang lebih baik, inilah diagram fish bone.

Tabel 1. Hasil Analisis Fishbone [1]

Faktor	Sub Masalah	Akibat	Penjelasan
Siswa	Pemanfaatan Passing Grade Belum Efisien	Pemanfaatan Passing Grade Belum Efisien	Siswa hanya berpatokan Pada Passinggrade yang berupa Nilai Rata Rata Terendah
	Asal Pilih Sekolah	Kehilangan Satu kesempatan sekolah negeri	Siswa memilih sekolah yang di inginkan tetapi tidak melihat nilai dan passing grade nya yang kurang effisien

Faktor	Sub Masalah	Akibat	Penjelasan
Sekolah	Sekolah Hanya Memberikan Data hanya satu tahun yang lalu Kurangnya mendapatkan informasi untuk calon peserta	Kurangnya mendapatkan informasi untuk calon peserta	Pemberitahuan data kurang maksimal sedangkang kan hanya 1 tahun kebelakang
	Data Kurang Valid	Informasi berupa data tidak optimal	
Algoritma	Melatih Algoritma	Masih Menggunakan patokan lama	Telalu banyak Algoritma yang di pakai dan atribut yang banyak dan sudah ada tetapi tidak di gunakan
	Penetapan Algoritma	Penetapan Algoritma yang belum sempurna	
Proses dan Data	Pengolahan Data yang kurang efisien	Data Yang dari kegiatan PPDB adalah mutlak	Siswa hanya berpatokan Pada Passing grade yang berupa Nilai Rata Rata Terendah
	Klasifikasi data kurang maksimal		Siswa memilih sekolah yang di inginkan tetapi tidak melihat nilai dan passing grade nya yang kurang efisien
	Banyak Atribut yang tidak di gunakan		

4.2 Data Yang di Gunakan

Contoh pembuatan tabel dapat dilihat pada Tabel 1 dan Tabel 2. Apabila isi tabel tersebut berasal dari sumber tertentu, tulis referensinya dengan angka [1].

Tabel 2. Data Arsip PPDB [2]

No.	NA	IND	MTK	ING	IPS	Sekolah
1	9.13	9.60	9.25	8.40	9.25	SMA 112
2	9.05	9.40	9.50	7.80	9.50	SMA 112
3	9.01	8.60	9.75	9.50	9.50	SMA 112
4	9.01	9.00	10.00	7.80	9.25	SMA 112
5	8.99	9.40	10.00	7.80	8.75	SMA 112
6	8.96	8.40	9.25	9.20	9.00	SMA 112
.....
2111	7.33	8.00	6.75	7.80	6.75	SMA 57
2112	7.33	7.40	8.00	7.40	6.50	SMA 57
2113	7.33	8.20	6.00	8.80	6.25	SMA 57

No.	NA	IND	MTK	ING	IPS	Sekolah
2114	7.31	8.60	5.75	7.60	7.25	SMA 57
2115	7.30	8.40	6.50	7.80	6.50	SMA 57
2116	7.30	7.80	6.50	8.40	6.50	SMA 57

Data yang digunakan dalam penelitian ini adalah sebanyak 2117 data dari arsip PPDB Sekolah Negeri di Jakarta Barat dan dengan sample 3 sekolah SMA Negeri 112, SMA Negeri 65, SMA Negeri 57, dari tahun 2013 sampai tahun 2018. Dengan atribut Nilai Akhir, Nilai Bahasa Indonesia, Nilai Matematika, Nilai Bahasa Inggris, Nilai IPA, Sekolah. Untuk attribut detailnya Nilai Akhir, Nilai Bahasa Indonesia, Nilai Matematika, Nilai Bahasa Inggris, Nilai IPA dengan bentuk nilai decimal dan untuk Sekolah dengan nilai bentuk label.

4.2 Perbandingan Setiap 3 Algoritma

4.2.1 Preprocessing

Preprocessing merupakan salah satu tahapan yang penting untuk data pada proses Machine Learning. Tahapan awal pada Perbandingan 3 Algoritma yaitu Normalisasi Standarisasi (Zero Mean). Normalisasi data digunakan untuk memodifikasi nilai dalam variabel sehingga dapat mengukur data ke dalam skala umum. Tujuan yang paling umum normalisasi adalah untuk mengubah nilai-nilai sehingga range nilainya berada diantara 0 dan 1. Transformasi nilai pada data menggunakan persamaan berikut :

$$\text{nilai baru} = \frac{(\text{nilai lama} - \text{nilai min}) \times (\text{range max} - \text{range min})}{(\text{nilai max} - \text{nilai min})} + \text{range min} \quad [1]$$

Dimana range min = 0, dan range max = 1, Sebagai contoh perhitungan matematis nya, menggunakan Nilai akhir 9.13 yang ada pada 4.1 Tabel Data dari Arsip PPDB.

$$\text{nilai baru} = \frac{(9.13 - 3.25) \times (1 - 0)}{(9.62 - 3.25)} + 0 \quad [2]$$

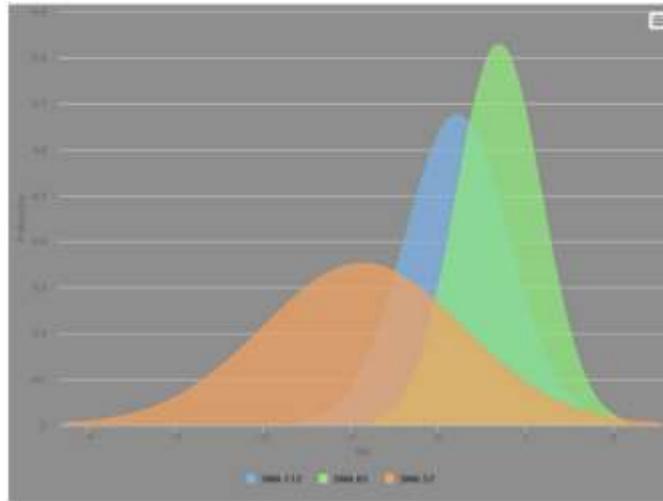
Tabel 3. Hasil Transformasi 0-1 [3]

No.	NA	IND	MTK	ING	IPS	Sekolah
1	9.13	9.60	9.25	8.40	9.25	SMA 112
2	9.05	9.40	9.50	7.80	9.50	SMA 112
3	9.01	8.60	9.75	9.50	9.50	SMA 112
4	9.01	9.00	10.00	7.80	9.25	SMA 112
5	8.99	9.40	10.00	7.80	8.75	SMA 112
.....
2111	7.33	8.00	6.75	7.80	6.75	SMA 57
2112	7.33	7.40	8.00	7.40	6.50	SMA 57
2113	7.33	8.20	6.00	8.80	6.25	SMA 57
2114	7.31	8.60	5.75	7.60	7.25	SMA 57
2115	7.30	8.40	6.50	7.80	6.50	SMA 57
2116	7.30	7.80	6.50	8.40	6.50	SMA 57

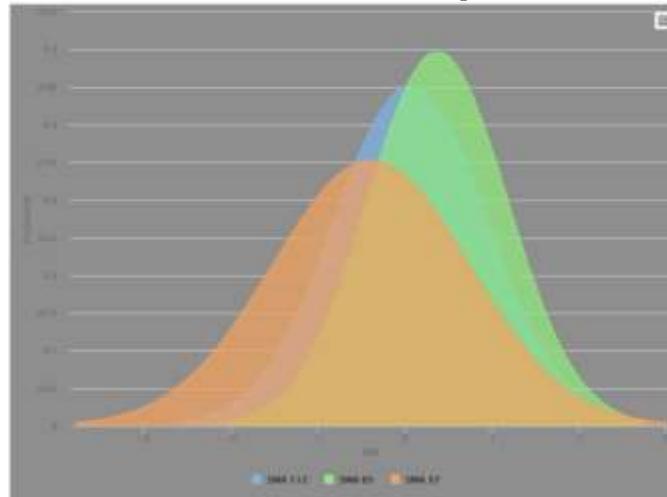
Pada Tabel 3, nilai pada setiap atribut telah diubah menjadi skala 0-1, sehingga dapat memudahkan pemrosesan yang akan dilakukan selanjutnya dan di visualisasikan. Data yang digunakan adalah data dari arsip ppdb dari tahun 2013 sampai dengan 2018 yang berjumlah 2216 data lihat table (Tabel 2. Sample Data PPDB), data dibagi menjadi dua untuk proses training sebanyak 1694 data, dan testing sebanyak 423 data. Sebelum dilakukan analisa, data dinormalkan dengan melakukan transformasi data menjadi skala 0-1 untuk memudahkan pemrosesan. Dari data uji yang ada, data cukup beragam dan menyebar, penyebaran data berdasarkan 5 variabel dapat dilihat pada visualisasi gambar (Gambar 1. Nilai Akhir Terhadap Sekolah), (Gambar 2. Nilai Bahasa Indonesia terhadap Sekolah), (Gambar 3. Nilai Matematika Terhadap Sekolah), (Gambar 4. Nilai Bahasa Inggris terhadap sekolah), (Gambar 5. Nilai Ilmu Pengetahuan Alam

JURNAL ILMIAH TEKNIK MESIN, ELEKTRO DAN KOMPUTER Vol.1, No.3, November 2021, pp. 69-86

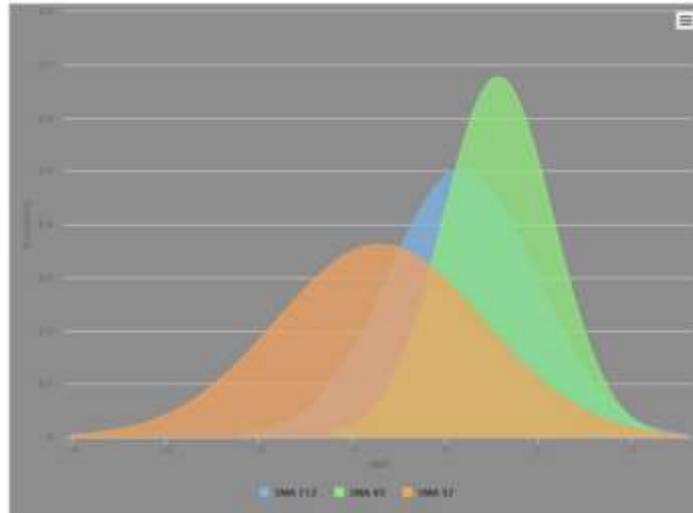
Terhadap Sekolah) . dari gambar ter visualisasikan data independen / yang mempengaruhi (Nilai Akhir, Nilai Bahasa Indonesia, Nilai Matematika, Nilai Bahasa Inggris, Nilai IPA) dan yang dipengaruhi atau dependen (Sekolah).



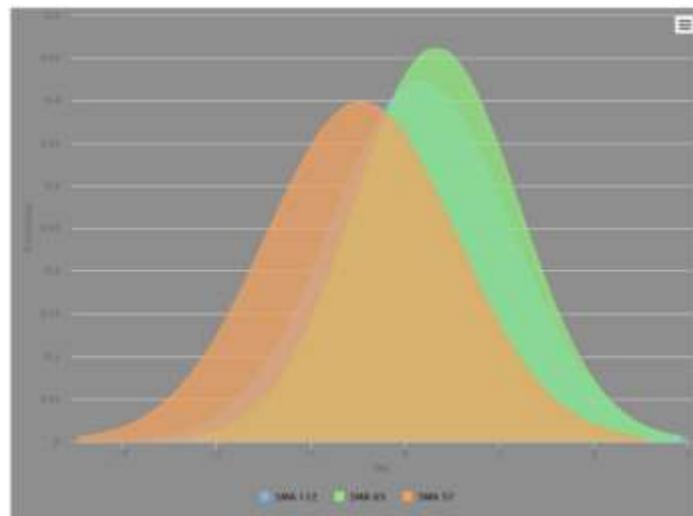
Gambar 3. Nilai Akhir Terhadap Sekolah



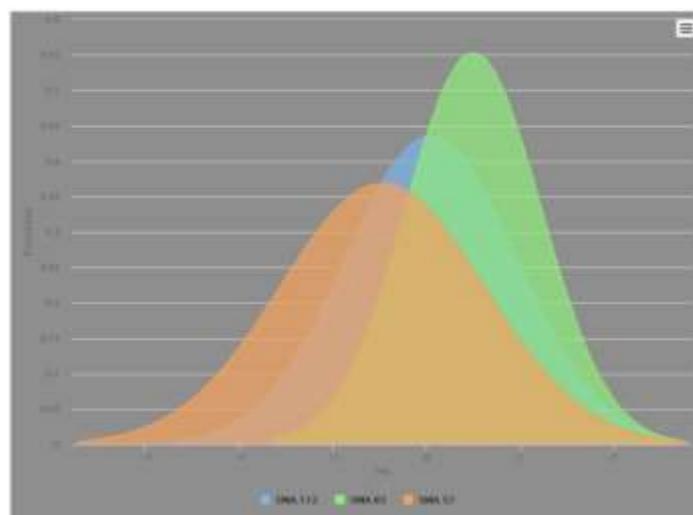
Gambar 4. Nilai Bahasa Indonesia terhadap Sekolah



Gambar 5. Nilai Matematika Terhadap Sekolah



Gambar 6. Nilai Bahasa Inggris terhadap sekolah



Gambar 7. Nilai Ilmu Pengetahuan Alam Terhadap Sekolah

4.2.2 Support Vector Machine

Klasifikasi menggunakan metode SVM adalah dengan mencari garis pemisah (hyperplane) antar hasil klasifikasi. Data yang digunakan adalah data ata arsip PPDB yang terdiri dari berbagai atribut polynominal dan numeric, serta data tersebut tidak linier, maka metode SVM yang digunakan adalah Kernel Radian Basis Function (RBF). Metode RBF ini memerlukan nilai parameter cost (C) dan gamma yang nilainya ditentukan oleh peneliti. Untuk mengklasifikasikan data yang tidak dapat dipisahkan secara linier formula SVM harus dimodifikasi karena tidak akan ada solusi yang ditemukan. Oleh karena itu, kedua bidang pembatas harus diubah sehingga lebih fleksibel (untuk kondisi tertentu) dengan penambahan variabel ξ_i menjadi $x_i \cdot w + b \geq 1 - \xi_i$ untuk kelas 1 dan $x_i \cdot w + b \leq -1 + \xi_i$ untuk kelas 2 dan seterusnya jika ada lebih dari satu class. Pencarian bidang pemisah terbaik dengan dengan penambahan variabel ξ_i sering juga disebut soft margin hyperplane. Dengan demikian formula pencarian bidang pemisah terbaik berubah menjadi:

$$K(x_i, x) = \exp(-\gamma |x_i - x|^2), \gamma > 0 \quad (1)$$

Oleh karena itu, peneliti di sini menggunakan uji coba nilai C sebanyak 4, yaitu 1, 3, 5, dan 10, serta uji coba nilai gamma sebanyak 4, yaitu 1, 2, 3, 4.

Berikut ini adalah nilai error dari masing-masing nilai C dan gamma untuk data PPDB untuk sample 3 sekolah;

Tabel 4. Tingkat Akurasi Model SVM Kernel RBF

SVM KERNEL RBF			
Cost	Gamma	Accuracy True	Accuracy False
1	1	64.07	35.93
1	3	62.88	37.12
1	5	63.36	36.64
1	10	63.12	36.88
2	1	63.36	36.64
2	3	64.54	35.46
2	5	63.83	36.17
2	10	63.59	36.41
3	1	63.36	36.64
3	3	64.78	35.22
3	5	64.07	35.93
3	10	62.17	37.83
4	1	63.83	36.17
4	3	64.54	35.46
4	5	62.17	37.83
4	10	60.52	39.48

Tabel di atas menunjukkan bahwa metode SVM menggunakan kernel RBF menghasilkan nilai error paling kecil untuk parameter C=3 dan gamma=3, sedangkan untuk hasil error paling besar adalah untuk C=10 dan gamma=4, sehingga nilai C dan gamma yang digunakan untuk nilai parameter di kernel RBF adalah yang menghasilkan error paling kecil, yaitu C=3 dan gamma=3, serta nilai akurasi untuk model pada data train adalah sebagai berikut;

Tabel 5. Hasil Prediksi Siswa menggunakan Algoritma Model Kernel RBF

	true SMA 112	true SMA 65	true SMA 57
pred. SMA 112	103	39	31
pred. SMA 65	33	77	4

PERBANDINGAN TIGA ALGORITMA CLASSIFIER UNTUK PENENTUAN PENERIMAAN PESERTA DIDIK BARU PADA SEKOLAH MENENGAH ATAS (Yulhendri)

Berdasarkan tabel di atas, banyaknya data Dari Arsip PPDB adalah sebanyak 2116 siswa yang di bagi menjadi data train 1694 siswa dan data test 423 siswa, hasil prediksi SVM benar untuk SMA 112 adalah sebanyak 103 siswa dan prediksi salah 63 Siswa . Kemudian benar untuk SMA 65 adalah sebanyak 77 siswa dan prediksi salah 51 Siswa. Dan benar untuk SMA 57 adalah sebanyak 94 siswa dan prediksi salah 35 Siswa.

$$\text{akurasi} = \frac{\text{jumlah data yang di prediksi benar}}{\text{jumlah data yang diuji}} \cdot 100\%$$

$$\begin{aligned} \text{akurasi SMA 112} &= \frac{103}{423} \cdot 100\% \\ &= 24.35\% \end{aligned}$$

$$\begin{aligned} \text{akurasi SMA 65} &= \frac{77}{423} \cdot 100\% \\ &= 18.20\% \end{aligned}$$

$$\begin{aligned} \text{akurasi SMA 57} &= \frac{94}{423} \cdot 100\% \\ &= 22.22\% \end{aligned}$$

$$\text{akurasi total} = 24.35\% + 18.20\% + 22.22\% = 64.78\%$$

Hasil akurasi untuk data PPDB dengan sample 3 sekolah diperoleh senilai 64.78%. Hasil akurasi sebesar 64.78% menunjukkan bahwa metode SVM dengan kernel RBF untuk melakukan prediksi terhadap data PPDB dengan sample 3 sekolah bisa dikatakan bisa memprediksi dengan bagus.

4.2.3 Naïve Bayes

Naïve Bayes Classifier merupakan metode klasifikasi yang bersumber pada teorema Bayes, yakni sebuah metode klasifikasi yang menggunakan metode probabilitas dan statistik untuk memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya. Keuntungan penggunaan adalah bahwa metoda ini hanya membutuhkan jumlah data pelatihan (data train) yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses klasifikasi. Karena yang diasumsikan sebagai variabel independent, maka hanya varians dari suatu variabel dalam sebuah kelas yang dibutuhkan untuk menentukan klasifikasi, bukan keseluruhan dari matriks kovarians.

Adapun alur dari metode Naive Bayes adalah sebagai berikut :

1. Menghitung jumlah kelas / label.
2. Menghitung Jumlah Kasus Per Kelas
3. Kalikan Semua Variable Kelas
4. Bandingkan Hasil Per Kelas

Untuk persamaan teorema Naïve Bayes dibawah ini:

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

Keterangan :

- x : Data dengan class yang belum diketahui
- c : Hipotesis data merupakan suatu class spesifik
- P(c|x) : Probabilitas hipotesis berdasar kondisi (posteriori probability)
- P(c) : Probabilitas hipotesis (prior probability)
- P(x|c) : Probabilitas berdasarkan kondisi pada hipotesis

$P(x)$: Probabilitas c

4.2.3.1 Transformasi Data

Sebelum implimentasi pada proses Naïve Bayes data diubah menjadi bentuk yang sesuai untuk diproses dalam Machine Learning Naïve Bayes. Dalam penelitian ini data yang akan diproses dari Microsoft Excel 2016 akan digunakan untuk pengolahan pada Software Rapid Miner. Dataset adalah data yang sudah ditransformasikan kedalam excel. Penulis melakukan transformasi dan inialisasi data agar penelitian ini bisa berjalan dengan baik. Data yang akan diinisialisasi merupakan data dari atribut Nilai Akhir, Nilai Bahasa Indonesia, Nilai Matematika, Nilai Bahasa Inggris, Nilai IPA seperti tabel berikut ini:

Tabel 6. Inialisasi Data

NA		IND		MTK		ING		IPA	
8.346 - 10	Diantara 8.346 - 10	9.120 - 10	Diantara 9.120 - 10	8.650 - 10	Diantara 8.650 - 10	8.760 - 10	Diantara 8.760 - 10	9 - 10	Diantara 9 - 10
7.072 - 8.346	Diantara 7.072 - 8.346	8.240 - 9.120	Diantara 8.240 - 9.120	7.300 - 8.650	Diantara 7.300 - 8.650	7.520 - 8.760	Diantara 7.520 - 8.760	9-9	Diantara 9-9
5.798 - 7.072	Diantara 5.798 - 7.072	7.360 - 8.240	Diantara 7.360 - 8.240	5.950 - 7.300	Diantara 5.950 - 7.300	6.280 - 7.520	Diantara 6.280 - 7.520	7-8	Diantara 7-8
4.524 - 5.798	Diantara 4.524 - 5.798	6.480 - 7.360	Diantara 6.480 - 7.360	4.600 - 5.950	Diantara 4.600 - 5.950	5.040 - 6.280	Diantara 5.040 - 6.280	6-7	Diantara 6-7
0 - 4.524	Diantara 0 - 4.524	0 - 6.480	Diantara 0 - 6.480	0 - 4.600	Diantara 0 - 4.600	0 - 5.040	Diantara 0 - 5.040	0-7	Diantara 0-7

Selanjutnya data yang sudah di inialisasi di transformasikan kedalam seperti table (Tabel 7. Hasil Inialisasi pada Data) dibawah ini hanya menggunakan 9 data dan selanjutnya.

Tabel 7. Hasil Inialisasi Data

Sekolah	NA	IND	MTK	ING	IPA
SMA 112	Lebih dari 8.35	Lebih dari 9.12	Lebih dari 8.65	Diantara 75.2 dan 8.76	Lebih dari 9
SMA 112	Lebih dari 8.35	Dintara dan 9.0	8.25	Lebih dari 8.65	Diantara 75.2 dan 8.76
SMA 112	Lebih dari 8.35	Dintara dan 9.0	8.25	Lebih dari 8.65	Diantara 75.2 dan 8.76
SMA 112	Lebih dari 8.35	Lebih dari 9.12	Lebih dari 8.65	Diantara 75.2 dan 8.76	Lebih dari 9
SMA 112	Lebih dari 8.35	Dintara dan 9.0	8.25	Lebih dari 8.65	Lebih dari 8.76
SMA 112	Lebih dari 8.35	Dintara dan 9.0	8.25	Lebih dari 8.65	Lebih dari 8.76
SMA 112	Lebih dari 8.35	Dintara dan 9.0	8.25	Lebih dari 8.65	Diantara 75.2 dan 8.76
SMA 112	Lebih dari 8.35	Lebih dari 9.12	Lebih dari 8.65	Diantara 75.2 dan 8.76	Lebih dari 9
SMA 112	Lebih dari 8.35	Dintara dan 8.3	7.35	Lebih dari 8.65	Diantara 75.2 dan 8.76

PERBANDINGAN TIGA ALGORITMA CLASSIFIER UNTUK PENENTUAN PENERIMAAN PESERTA DIDIK BARU PADA SEKOLAH MENENGAH ATAS (Yulhendri)

Sekolah	NA	IND	MTK	ING	IPA
.....

4.2.3.2 Hasil Pengujian Akurasi

Pada tahap ini dilakukan pemodelan data, metode yang dipakai pada penelitian ini adalah probabilitas (prediksi) dengan menggunakan algoritma Naive Bayes. Data yang telah dikumpulkan, dan ditransformasi akan dikelola menggunakan probabilitas untuk melihat performa tingkat akurasi Algoritma Naive Bayes.

Untuk menghitung akurasi Naive Bayes algoritma sebagai berikut

Tabel 8. Hasil Prediksi Siswa menggunakan Algoritma Naive Bayes

	true SMA 112	true SMA 65	true SMA 57
pred. SMA 112	61	36	54
pred. SMA 65	73	80	1
pred. SMA 57	34	1	83

Jumlah data yang diuji hasil prediksi SVM benar untuk SMA 112 adalah sebanyak 61 siswa dan prediksi salah 107 Siswa. untuk SMA 65 adalah sebanyak 80 siswa dan prediksi salah 37 Siswa. Dan untuk SMA 57 adalah sebanyak 83 siswa dan prediksi salah 55 Siswa.

$$\text{akurasi} = \frac{\text{jumlah data yang di prediksi benar}}{\text{jumlah data yang diuji}} \cdot 100\%$$

$$\begin{aligned} \text{akurasi SMA 112} &= \frac{61}{423} \cdot 100\% \\ &= 14.42\% \end{aligned}$$

$$\begin{aligned} \text{akurasi SMA 65} &= \frac{80}{423} \cdot 100\% \\ &= 18.91\% \end{aligned}$$

$$\begin{aligned} \text{akurasi SMA 57} &= \frac{83}{423} \cdot 100\% \\ &= 19.61\% \end{aligned}$$

$$\begin{aligned} \text{akurasi total} &= 14.42\% + 18.91\% + 19.61\% \\ &= 52.94\% \end{aligned}$$

Hasil akurasi untuk data PPDB dengan sample 3 sekolah diperoleh senilai 52.94%. Hasil akurasi sebesar 52.94% menunjukkan bahwa metode Naive Bayes untuk melakukan prediksi terhadap data PPDB dengan sample 3 sekolah kurang bisa dikatakan bisa memprediksi dengan bagus.

4.2.4 Decesion Tree

Pohon Keputusan atau dikenal dengan Decision Tree adalah salah satu metode klasifikasi yang menggunakan representasi suatu struktur pohon yang berisi alternatif-alternatif untuk pemecahan suatu masalah. Pohon ini juga menunjukkan faktor-faktor yang mempengaruhi hasil alternatif dari keputusan tersebut disertai dengan estimasi hasil akhir bila kita mengambil keputusan tersebut.

1. Root node, merupakan node paling atas (akar) dimana pada node tidak ada input dan mempunyai output lebih dari satu. Simpul akar biasanya berupa atribut yang paling memiliki pengaruh terbesar pada suatu kelas tertentu.
2. Internal node, merupakan node percabangan dimana pada internal node hanya terdapat satu input dan mal 2 output (berderajat $\neq 0$).

- Leaf node atau terminal node, merupakan node akhir dimana pada node hanya terdapat satu input dan tidak mempunyai output (berderajat 0).

Dasar dari algoritma C4.5 adalah pembentukan pohon keputusan (decision tree), dimana cabang-cabang pohon keputusan merupakan pertanyaan klasifikasi dan daun-daunnya merupakan kelas atau segmennya. Secara umum algoritma C4.5 untuk membangun Decision Tree dilakukan dengan beberapa tahapan sebagai berikut:

- Pilih atribut sebagai akar.
- Buat cabang untuk tiap-tiap nilai.
- Buat kasus dalam cabang.
- Ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama.

Untuk menghitung nilai gain, maka terlebih dahulu perlu menghitung nilai Entropy. Entropy merupakan distribusi probabilitas dalam teori informasi dan diadopsi kedalam algoritma C4.5 untuk mengukur tingkat homogenitas distribusi kelas dari sebuah himpunan (dataset). Sebagai ilustrasi, semakin tingkat entropy dari sebuah dataset maka semakin homogen distribusi kelas pada dataset tersebut.

Untuk persamaan Algoritma Decision tree

$$Entropy(S) = \sum_{i=1}^n -p_i \cdot \frac{\ln(p_i)}{\ln(2)}$$

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{S_i}{S} \cdot Entropy(S_i)$$

Dimana :

S = Himpunan kasus

n = Jumlah partisi S

pi = Probabilitas yang didapat dari (Ya/Tidak) dibagi total kasus

A = Atribut N = Jumlah partisi atribut A

|Si| = Jumlah kasus pada partisi ke-i

|S| = Jumlah kasus dalam S

4.2.4.1 Menentukan Akar Dari Pohon Keputusan

Dari data sebanyak 1694 siswa dan data testing sebanyak 423 siswa yang diperoleh dari dataset Arsip PPDB 3 sekolah (Tabel 4.1 Data dari Arsip PPDB). Yang akan dihitung nilai entropy, information gain, dan gain ratio yang memiliki fitur berjumlah 3 menggunakan tipe polinomial dengan rincian sebagai berikut :

- SMA 112= kontinu
- SMA 65 = kontinu
- SMA 57 = kontinu

Simpul Akar		Total Siswa		Range Nilai				Entropi Total
		2116	0 – 5.9	6 – 6.9	7 – 7.9	8 – 8.9	9 - 10	
Nilai Akhir	SMA 112	842	0	0	65	737	40	
	SMA 65	586	0	0	0	430	156	
	SMA 57	689	16	33	205	432	3	
	Entropi tiap kelas		0.05329	0.09362	0.37901	0.30542	0.32074	
	Gain				1.25791			
Nilai	SMA 112	842	0	3	58	486	295	
	SMA 65	586	0	1	17	281	287	

PERBANDINGAN TIGA ALGORITMA CLASSIFIER UNTUK PENENTUAN PENERIMAAN PESERTA DIDIK BARU PADA SEKOLAH MENENGAH ATAS (Yulhendri)

Bahasa Indonesia	SMA 57	689	1	14	131	378	165	
	Entropi tiap kelas		0.00522	0.05850	0.32717	0.47942	0.53030	
	Gain		0.47898					
Nilai Matematika	SMA 112	842	4	15	58	306	429	1.56910
	SMA 65	586	0	1	17	136	428	
	SMA 57	689	34	85	131	238	140	
	Entropi tiap kelas		0.47898	0.20949	0.40022	0.52630	0.51154	
	Gain		0.47896					
Nilai Bahasa Inggris	SMA 112	842	9	15	200	370	215	
	SMA 65	586	1	1	112	269	181	
	SMA 57	689	18	85	262	251	53	
	Entropi tiap kelas		0.08257	0.29841	0.51058	0.52553	0.47458	
	Gain		0.47896					
Nilai Ilmu Pengetahuan Alam	SMA 112	842	2	33	184	419	204	
	SMA 65	586	0	0	51	297	238	
	SMA 57	689	13	76	225	286	89	
	Entropi tiap kelas		0.05062	0.22042	0.47862	0.51069	0.50052	
	Gain		0.48319					

4.2.4.2 Hasil Pengujian Akurasi

Uji coba ini dilakukan untuk mengetahui apakah sistem yang dibuat telah sesuai untuk menilai akurasi prediksi data PPDB 3 sekolah dengan Support Vector Machine. Uji coba dilakukan dengan menentukan data testing PPDB yang dipilih secara random (). Data testing tersebut akan dicari nilai kelayakannya menggunakan aplikasi pembandingan berupa Rapidminer 9.6.

Hasil analisa menggunakan sistem dan rapidminer dapat dilihat di Gambar 4.6 Perhitungan Akurasi Prediksi metode Decesion Tree. Untuk menghitung akurasinya sebagai berikut:

Tabel 9. Hasil Prediksi Data PPDB 3 Sekolah Algoritma Decesion Tree

	true SMA 112	true SMA 65	true SMA 57
pred. SMA 112	84	52	8
pred. SMA 65	31	64	8
pred. SMA 57	51	12	113

$$akursi = \frac{\text{jumlah data yang di prediksi benar}}{\text{jumlah data yang diuji}} \cdot 100\%$$

$$akurasi SMA 112 = \frac{84}{423} \cdot 100\% \\ = 19.86\%$$

$$akurasi SMA 65 = \frac{64}{423} \cdot 100\% \\ = 15.13\%$$

$$akurasi SMA 57 = \frac{113}{423} \cdot 100\% \\ = 26.71\%$$

$$akurasi total = 19.86\% + 15.13\% + 26.71\%$$

= 61.7%

Hasil akurasi untuk data PPDB dengan sample 3 sekolah diperoleh senilai 61.7%. Hasil akurasi sebesar 61.7% menunjukkan bahwa Algoritma Naive Bayes untuk melakukan prediksi terhadap data PPDB dengan sample 3 sekolah dikatakan bisa memprediksi lumayan bagus.

4.3 Hasil Perbandingan 3 Algoritma

Setelah dilakukan analisis klasifikasi menggunakan SVM kernel RBF, Naive Bayes, Dan Decesion Tree mendapatkan hasil nilai akurasi dari masing-masing menggunakan data dari arsip PPDB selama 5 tahun dengan 3 sekolah sebagai data set. Berikut ini adalah perbandingan tingkat akurasi dari Algoritma SVM, Naive Bayes, dan Decesion Tree dengan data set yang di bagi 80% sebagai data train dan 20% data test. Yang nanti di implementasi kan dengan Algoritma yang mempunyai tingkat akurasi tertinggi.

Support Vector Machine	Naive Bayes	Decesion Tree
64.78 %	52.94%	61.7%

Tabel 4.11 di atas, menjelaskan tentang perbandinagn tingkat akurasi dari Algoritma SVM , Naive Bayes , dan Decesion Tree mendapatkan bawa data terbaik adalah hasil Klasifikasi data Sekolah pada kegiatan PPDB megggunakan Algoritma SVM dengan gamma=3 dan Cost =3 dan nilai akurasi sebesar 64.78 %. Algoritma SVM dengan nilai akurasi tertinggi dan akan di implementasi kan pada Sistem Prediksi Sekolah Menengah Atas PPDB.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil yang telah diuraikan, dapat ditarik beberapa kesimpulan, diantaranya :

1. Sistem Machine Leraning menggunakan Algoritma SVM dengan nilai akurasi terbaik dari 2116 siswa dari Arsip PPDB, Dan di buatnya model Machine Learning dan di implementasikan Web App.
2. Metode Machine Learning yang menggunakan model Algoritma SVM yang akan menampilkan hasil prediksi sekolah pada kegiatan ppdb menggunakan nilai dari hasil tryout bimbingan belajar dan hasil nilai ujian Nasional.
3. Metode Machine Learning SVM ini dapat mengantisipasi kegagalan calon peserta dalam memilih sekolah pada Kegiatan PPDB.

Saran yang berkaitan dengan pengembangan Sistem Machine Learning ini :

1. Diharapkan tsistem ini dapat dikembangkan lebih jauh lagi agar masyarakat semakin dengan mudah lagi dalam memahami fitur yang ada pada sistem ini.
2. Informasi yang ada dalam sistem Machine Learning ini pun diharapkan dapat dikembangkan lagi, agar calon peserta PPDB yang menggunakan sistem ini dapat mengetahui prediksi sekolah menggunakan nilai tryout bimbingan belajar dan nilai ujian akhir.

Ucapan Terima Kasih

Jika perlu berterima kasih kepada Dekan Fasilkom Universitas Esa Unggul yang telah mendukung proses penelitian ini.

DAFTAR PUSTAKA

Handriani, I. et al. "Pengaruh Aktivitas Fisik Jalan Pagi Terhadap Tekanan Darah Pada Lansia Dengan Penyakit Hipertensi Di Upt Pstw Khususul Khotimah Pekanbaru Endi", <http://jurnal.ensiklopediaku.org> Ensiklopedia of Journal, 'Vol. 1 No.2 Edisi 1 Januari 2019'

Harrington, P. (2015) Machine Learning in Action, Efficient Learning Machines. doi: 10.1007/s10994-011-5249-4.

Khodijah, S. and Labibah, L. (2019) 'TEACHING READING USING TALKING STICK', PROJECT (Professional Journal of English Education). doi: 10.22460/project.v2i1.p40-45.

Larose, D. T. (2005) *Discovering Knowledge in Data: An Introduction to Data Mining*, Discovering Knowledge in Data: An Introduction to Data Mining. doi: 10.1002/0471687545.

Maimon, O. and Rokach, L. (2010) *Data Mining and Knowledge Discovery Handbook 2ed*, Springer. doi: 10.1007/978-0-387-09823-4.

Pressman, R. S. (2010) *Rekayasa Perangkat Lunak: Pendekatan Praktisi, Software Engineering A Practitioner's Approach 7th Ed* - Roger S. Pressman. doi: 10.1017/CBO9781107415324.004.

Rahmat C.T.I., B. et al. (2017) 'Implementasi k-means clustering pada rapidminer untuk analisis daerah rawan kecelakaan', Seminar Nasional Riset Kuantitatif Terapan 2017.

Setiawan, R. (2016) 'Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Promosi Mahasiswa Baru (Studi Kasus : Politeknik Lp3i Jakarta)', J. Lentera Ict.

Sunjana (2010) 'Aplikasi Mining Data Mahasiswa dengan Metode Klasifikasi Decision Tree', Seminar Nasional Aplikasi Teknologi Informasi. doi: 10.1023/B:CRIS.0000041037.22694.0b.

Witten, I. H. et al. (1999) 'Weka: Practical Machine Learning tools and techniques with Java implementations', Proc ICONIP/ANZIIS/ANNES'99 Int. Workshop: Emerging Knowledge Engineering and Connectionist-Based Info. Systems. doi: 10.1.1.16.949.

Yuniarto, E. A. (2014) 'Perbaikan pada fishbone diagram sebagai root cause analysis tool', Jurnal Teknik Industri, pp. 217–224.