

PERBANDINGAN ALGORITMA NAÏVE BAYES , K-NN , ID3 , DAN SVM DALAM MENENTUKAN PREDIKSI KELULUSAN SISWA DI SMK MUHAMADIAH MAJENANG

¹Hani Latifah, ²Sri Mujiyono

^{1,2}Univesitas Ngudi Waluyo

Email: ¹hanilatifahh@gmail.com, ²Sri Mujiyono

Abstrak

SMK Muhamadiah Majenang merupakan unit pelaksana pendidikan formal untuk menyiapkan peserta didik agar bisa menjadi generasi penerus bangsa. Kelulusan menjadi syarat dan ketentuan yang di tetapkan agar bisa menyelesaikan proses pembelajaran yang telah di tentukan dan harus di lalui oleh setiap siswa. Memprediksi tingkat kelulusan siswa penting bagi penyelenggara pendidikan untuk meningkatkan dan mempertahankan prestasi siswa dalam proses pembelajaran. Maka dengan melakukan penelitian prediksi kelulusan siswa menggunakan Data Mining diharapkan dapat menjadi upaya dalam peningkatan kualitas pendidikan. Dalam penelitian ini akan di lakukan perbandingan antara algoritma naïve bayes , K-NN , ID3 , dan SVM untuk menentukan metode mana yang paling efektif dalam menentukan prediksi kelulusan siswa. Atribut yang digunakan dalam penelitian ini adalah Transkrip Nilai dari Semester 1 sampai Semester 5 dan Ujian Sekolah. Dalam penelitian ini menggunakan populasi 420 siswa yang terdiri dari 326 siswa laki-laki dan 94 siswa perempuan.

Kata Kunci : *Kelulusan Siswa , Algoritma Naïve Bayes , K-NN , ID3 , SVM*

Abstract

SMK Muhamadiah Majenang is a formal education implementing unit to prepare students to become the nation's next generation. Graduation is the terms and conditions that are set in order to complete the learning process that has been determined and must be passed by each student. Predicting student graduation rates is important for education providers to improve and maintain student achievement in the learning process. So by conducting research on student graduation predictions using Data Mining, it is hoped that it can be an effort to improve the quality of education. In this study, a comparison will be made between the nave Bayes algorithm, K-NN, ID3, and SVM to determine which method is the most effective in determining student graduation predictions. The attributes used in this study are transcripts of grades from semester 1 to semester 5 and school exams. In this study, a population of 420 students consisted of 326 male students and 94 female students.

Keywords: *Student Graduation, Naïve Bayes Algorithm, K-NN, ID3, SVM*

PENDAHULUAN

Pada dasarnya kelulusan siswa dipengaruhi oleh kemampuan siswa dalam memenuhi syarat kelulusan yang ditetapkan dalam rapat kelulusan yang di tanda tangani oleh Kepala Sekolah melalui surat

keputusan dari hasil rapat. mengenai tingkat kelulusan digunakan untuk menentukan strategi dan atau membuat kebijakan baru sehingga meningkatkan kelulusan pada tahun berikutnya bisa diketahui dari data tersebut.

Setiap siswa memiliki tingkat motivasi, sikap, dan respon yang berbeda pada saat proses belajar mengajar. Ada siswa yang aktif dan rajin pada saat proses pembelajaran dan ada yang kurang aktif dalam menanggapi apa yang di sampaikan pada saat proses belajar. Hal itu dapat berpengaruh terhadap hasil ujian-ujian siswa, sehingga ada siswa yang langsung dapat menyelesaikan pendidikannya dan ada yang harus di berikan tugas tambahan agar dapat menyelesaikan pendidikannya.

Prediksi dilakukan dengan memanfaatkan data-data yang sebelumnya dan dapat membandingkan antara Transkrip nilai dari Semester 1 sampai Semester 5 dan US (Ujian Sekolah) yang dapat menentukan faktor mana saja yang menjadi penunjang yang paling dominan didalam salah satu syarat tingkat kelulusan seorang siswa.

Teknik data mining dan machine learning dapat digunakan untuk memprediksi berdasarkan data-data masa lalu. Data mining adalah proses untuk menemukan pola yang berguna dan kecenderungan di dalam kumpulan data yang besar. Dari sumber lain, data mining adalah ilmu yang mempelajari tentang pengumpulan, pembersihan, pengolahan, analisis, dan memperoleh wawasan yang berguna dari data (Aggarwal, 2015, p. 1). Salah satu metode yang dapat digunakan untuk memprediksi adalah klasifikasi.

Tugas dari klasifikasi adalah memprediksi keluaran variabel/class yang bernilai kategorikal atau polinomial (Kotu & Deshpande, 2015, p. 9).

Untuk melakukan tingkat kelulusan siswa dalam suatu ajaran sekolah setiap tahunnya dapat dilakukan suatu prediksi atau klasifikasi Model untuk prediksi yang banyak dilakukan oleh para peneliti diantaranya Naïve Bayes (NB), Neural Network (NN), Iterative Dichotomiser 3 (ID3) dan Support Vector Machine (SVM). Penulis akan membandingkan antara metode Naïve Bayes, K-NN, ID3, dan SVM dalam memprediksi siswa pada saat hasil kelulusan mereka.

METODE PENELITIAN

Hasil dari tahapan ini adalah menyiapkan perangkat keras yang digunakan untuk pembuatan desain sistem, pengembangan sistem, serta pengujiannya.

Perangkat pengembangan yang digunakan adalah :

- 1) Perangkat Keras
 - a. Laptop Acer
 - b. Jaringan Internet /WIFI
- 2) Perangkat Lunak

Data Mining : Rapid Miner

Metode yang dilakukan dalam mengumpulkan data, yaitu sebagai berikut: Metode yang digunakan untuk mengumpulkan data dalam penelitian ini adalah sebagai berikut:

1. Observasi
Dalam penyusunan skripsi ini, penulis melakukan observasi dengan salah satu staff guru yang ada di sekolah menggunakan WhatsApp bertujuan untuk mengumpulkan data-data siswa yang akan dijadikan populasi dalam tingkat kelulusan.
2. Wawancara
Selain observasi, penulis juga melakukan wawancara secara langsung dengan staff guru secara langsung untuk mendapatkan kelengkapan data yang dibutuhkan
3. Studi Pustaka
Penelitian ini melakukan studi pustaka dengan cara mempelajari dari berbagai macam sumber jurnal dan artikel yang berhubungan dengan pembahasan yang diperlukan guna untuk dijadikan sebagai referensi penelitian ini

Dalam metode ini digunakan untuk mengumpulkan dan mencari data tentang informasi yang dibutuhkan untuk melakukan penelitian. Sumber yang digunakan dalam penelitian ini yaitu:

1. Data Primer
Data yang diperoleh secara langsung dari sumber data yang berhubungan dengan penelitian. Data ini didapat dari data Nilai Transkrip Semester 1 sampai 5, Nilai US (Ujian Sekolah). Untuk mendapatkan data primer dapat

dilakukan dengan cara Observasi dan Wawancara secara langsung kepada narasumber dengan salah satu staff guru yang ada di sekolah yaitu Bapak Mukhlis.

2. Data Sekunder

Data sekunder adalah pelengkap dari data primer yang diperoleh dari buku-buku dan jurnal yang ada di internet.

Beberapa hal penting yang menjadi instrumen pada penelitian ini sebagai berikut:

1. Penelitian ini menggunakan data riset, data yang akan dimasukan dan dilakukan pengolahan yaitu data (Nilai Transkrip semester 1 sampai 5 dan Nilai US (Ujian Sekolah)) yang akan menjadi sampel sama dengan populasi digunakan sebagai instrumen guna memperoleh data dalam proses klasifikasi prediksi dalam tingkat kelulusan siswa.
2. Data yang sudah tersedia didalam bentuk Microsoft Excel adalah data sampel sama dengan populasi dengan keseluruhan siswa sebanyak 420 dengan siswa yang lulus maupun siswa yang lulus terlambat yang diambil dari tahun ajaran 2021
3. Perangkat lunak yang digunakan dalam pengolahan data adalah Microsoft Excel 2010.
4. Untuk melihat hasil akhir dari algoritma dan metode Naïve Bayes , K-NN , Decision Tree , SVM yang sudah digunakan untuk mengolah data yaitu menggunakan aplikasi Rapid Miner.

HASIL DAN PEMBAHASAN

1. Validasi dan Evaluasi

1) Validasi

Tujuan utama penelitian ini adalah untuk mengetahui nilai akurasi dari algoritma Naïve Bayes , K-NN , Decision Tree , dan SVM pada pembobotan atribut yang digunakan untuk memprediksi kelulusan siswa. Kemudian membandingkan keempat algoritma tersebut sehingga dapat

diperoleh salah satu algoritma yang terbaik

2) Evaluasi

Dalam tahap ini dilakukan pengukuran keakuratan hasil yang dicapai oleh model menggunakan beberapa teknik yang terdapat dalam framework RapidMiner versi 5.3 yaitu *confusion matrix* untuk pengukuran akurasi model, dan X-validation untuk validasi.

2. Metode Naïve Bayes.

1) Perhitungan manual

Tabel 1. Confusion Matrix Naive Bayes

	True lulus	True lulus terlambat
Pred.lulus	368	0
Pred.lulus terlambat	34	18

Dari tabel tersebut dapat dilakukan perhitungan :

- Accuracy

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} = \frac{368+18}{368+18+34+0} = \frac{386}{420} = 0,919 = 91,90\%$$
- Sensitivity

$$Sensitivity = \frac{TP}{TP+FN} = \frac{368}{368+0} = 1 = 100\%$$
- Specificity

$$Specificity = \frac{TN}{TN+FP} = \frac{18}{18+34} = 0,3462 = 34,62\%$$
- PPV

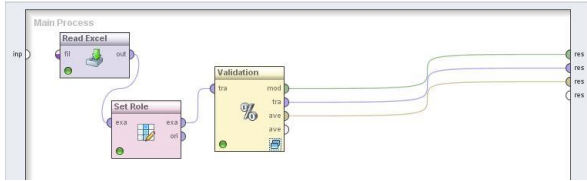
$$PPV = \frac{TP}{TP+FP} = \frac{368}{368+34} = 0,9154 = 91,54\%$$
- NPV

$$NPV = \frac{TN}{TN+FN} = \frac{18}{18+0} = 1 = 100\%$$

2) Pengujian dengan RapidMiner

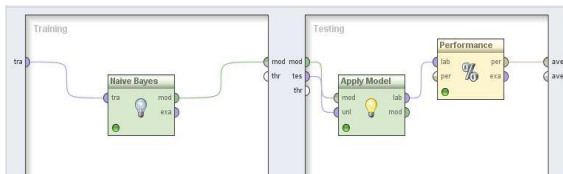
Dataset yang disimpan dalam format excel dibuka menggunakan operator Read Exel. Kemudian keluarannya dihubungkan ke Set Role terlebih dahulu untuk menentukan atribut label kemudian baru di hubungkan ke Validation (X-Validation).

Susunan operator ditunjukkan pada Gambar



Gambar 1. Process Naive Bayes

Untuk melatih dan menguji model di bagian training diisi operator Naive Bayes, pada testing diisi operator Apply Model dan Performance. Susunan operator training dan testing ditunjukkan pada Gambar



Gambar 2. Validasi Naive Bayes

Kemudian model dijalankan sehingga didapat hasil kinerja model seperti pada Gambar didapat akurasi 91,90 %.

	true lulus	true lulus terlambat	class precision
pred. lulus	368	0	100.00%
pred. lulus terlambat	34	18	34.62%
class recall	91.54%	100.00%	

Gambar 3. Akurasi Naive Bayes

3. Metode K-NN

1) Perhitungan Manual

Tabel 2. Confusion Matrix K-NN

	True lulus	True lulus terlambat
Pred.lulus	400	6
Pred.lulus terlambat	2	12

Dari tabel tersebut dapat dilakukan perhitungan :

- Accuracy

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} = \frac{400+12}{400+12+2+6} = \frac{412}{420} = 0,981 = 98,10\%$$
- Sensitivity

$$Sensitivity = \frac{TP}{TP+FN} = \frac{400}{400+6} = 0,9852 = 98,52\%$$
- Specificity

$$Specificity = \frac{TN}{TN+FP} = \frac{12}{12+2} = 0,8571 = 85,71\%$$

$$Specificity = \frac{TN}{TN+FP} = \frac{12}{12+2} = 0,8571 = 85,71\%$$

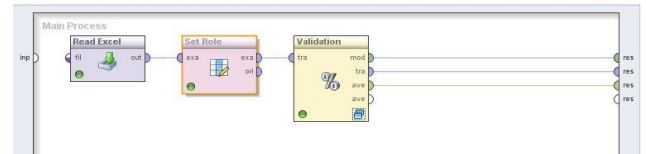
- PPV

$$PPV = \frac{TP}{TP+FP} = \frac{400}{400+2} = 0,995 = 99,50\%$$
- NPV

$$NPV = \frac{TN}{TN+FN} = \frac{12}{12+6} = 0,6667 = 66,67\%$$

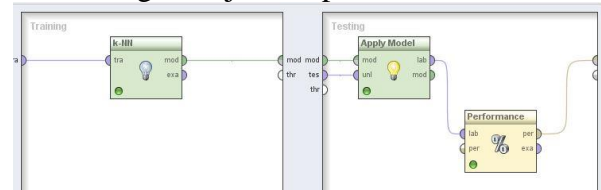
2) Pengujian dengan RapidMiner

Dataset yang disimpan dalam format excel dibuka menggunakan operator Read Excel. Kemudian keluarannya dihubungkan ke Set Role terlebih dahulu untuk menentukan atribut label kemudian baru di hubungkan ke Validation (X-Validation). Susunan operator ditunjukkan pada Gambar



Gambar 4. Process K-NN

Untuk melatih dan menguji model di bagian training diisi operator K-NN , pada testing diisi operator Apply Model dan Performance. Susunan operator training dan testing ditunjukkan pada Gambar



Gambar 5. Validasi K-NN

Kemudian model dijalankan sehingga didapat hasil kinerja model seperti pada Gambar didapat akurasi 98,10%.

	true lulus	true lulus terlambat	class precision
pred. lulus	400	6	98.52%
pred. lulus terlambat	2	12	85.71%
class recall	99.50%	66.67%	

Gambar 6. Akurasi K-NN

4. Metode Decision Tree (ID3)

1) Perhitungan Manual

Tabel 3. Confusion Matrix ID3

	True lulus	True lulus terlambat

Pred.lulus	402	4
Pred.lulus terlambat	0	14

Dari tabel tersebut dapat dilakukan perhitungan :

- Accuracy

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} = \frac{402+14}{402+14+0+4} = \frac{416}{420} = 0,9905 = 99,05\%$$
- Sensitivity

$$Sensitivity = \frac{TP}{TP+FN} = \frac{402}{402+4} = 0,9901 = 99,01\%$$
- Specificity

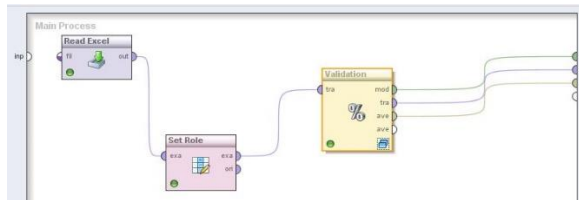
$$Specificity = \frac{TN}{TN+FP} = \frac{14}{14+0} = 1 = 100\%$$
- PPV

$$PPV = \frac{TP}{TP+FP} = \frac{402}{402+0} = 1 = 100\%$$
- NPV

$$NPV = \frac{TN}{TN+FN} = \frac{14}{14+4} = 0,7778 = 77,78\%$$

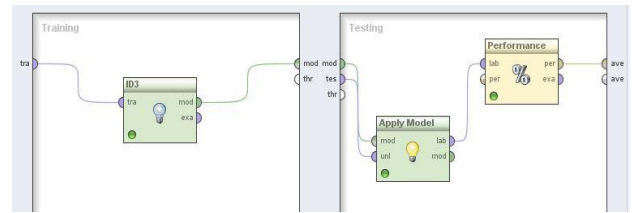
2) Pengujian dengan RapidMiner

Dataset yang disimpan dalam format excel dibuka menggunakan operator Read Excel. Kemudian keluarannya dihubungkan ke Set Role terlebih dahulu untuk menentukan atribut label kemudian baru di hubungkan ke Validation (X-Validation). Susunan operator ditunjukkan pada Gambar



Gambar 7. Process ID3

Untuk melatih dan menguji model di bagian training diisi operator ID3 , pada testing diisi operator Apply Model dan Performance. Susunan operator training dan testing ditunjukkan pada Gambar

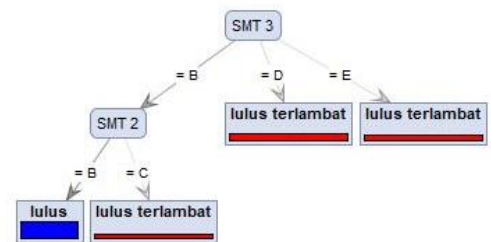


Gambar 8. Validasi ID3

Kemudian model dijalankan sehingga didapat hasil kinerja model seperti pada Gambar didapat akurasinya 99,05%.

	true lulus	true lulus terlambat	class precision
pred. lulus	402	4	99.01%
pred. lulus terlambat	0	14	100.00%
class recall	100.00%	77.78%	

Gambar 9. Akurasi ID3



Gambar 10. Pohon Keputusan ID3

4.1.4 Metode SVM

1) Perhitungan Manual

Tabel 4. Confusion Matrix SVM

	True lulus	True lulus terlambat
Pred.lulus	402	18
Pred.lulus terlambat	0	0

Dari tabel tersebut dapat dilakukan perhitungan :

- Accuracy

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} = \frac{402+0}{402+0+0+18} = \frac{402}{420} = 0,95,71 = 95,71\%$$
- Sensitivity

$$Sensitivity = \frac{TP}{TP+FN} = \frac{402}{402+18} = 0,9571 = 95,71\%$$
- Specificity

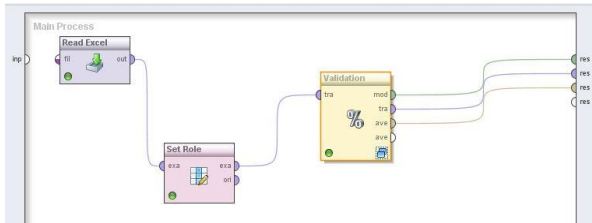
$$Specificity = \frac{TN}{TN+FP} = \frac{0}{0+0} = 0 = 0\%$$
- PPV

$$PPV = \frac{TP}{TP+FP} = \frac{402}{402+0} = 1 = 100\%$$
- NPV

$$NPV = \frac{TN}{TN+FN} = \frac{0}{0+18} = 0 = 0\%$$

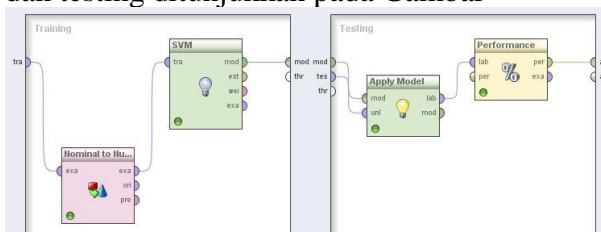
2) Pengujian dengan RapidMiner

Dataset yang disimpan dalam format excel dibuka menggunakan operator Read Excel. Kemudian keluarannya dihubungkan ke Set Role terlebih dahulu untuk menentukan atribut label kemudian baru di hubungkan ke Validation (X-Validation). Susunan operator ditunjukkan pada Gambar



Gambar 11. Process SVM

Untuk melatih dan menguji model di bagian training diisi operator SVM , Karena Linear Regression tidak mendukung masukan nominal, maka dikonversi menggunakan Nominal to Numeric. Kemudian keluarannya dihubungkan ke operator SVM , pada testing diisi operator Apply Model dan Performance. Susunan operator training dan testing ditunjukkan pada Gambar



Gambar 12 Validasi SVM

Kemudian model dijalankan sehingga didapat hasil kinerja model seperti pada Gambar didapat akurasi nya 95,71%.

Multiclass Classification Performance			
accuracy: 95.71% +/- 0.95% (mikro: 95.71%)			
	true lulus	true lulus terlambat	class precision
pred. lulus	402	18	95.71%
pred. lulus terlambat	0	0	0.00%
class recall	100.00%	0.00%	

Gambar 13. Akurasi SVM

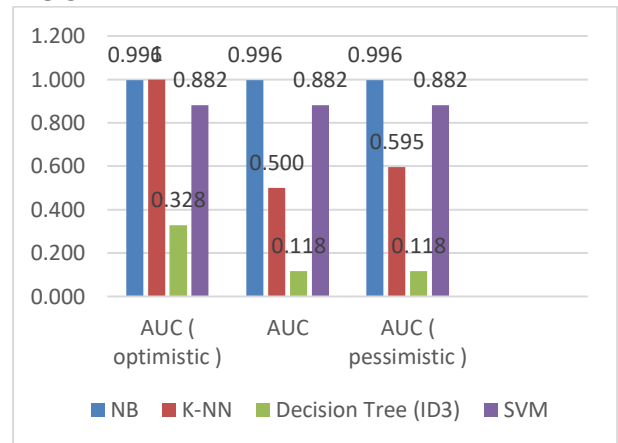
4.2 Hasil percobaan dan pengujian metode

Tabel di bawah ini memperlihatkan perbandingan akurasi dan AUC antara algoritma Naïve Bayes , K-NN , Decision Tree (ID3) , dan SVM.

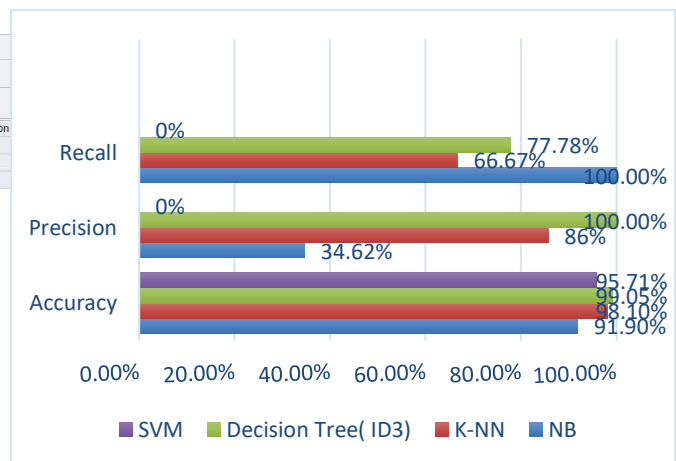
Gambar 14. Grafik perbandingan akurasi dan AUC antara algoritma Naïve Bayes , K-NN , Decision Tree (ID3) , dan SVM.

Grafik diatas menunjukkan Naïve Bayes 91,90 % , K-NN 98,10 % , Decision Tree(ID3) 99,05 % , dan SVM 95,71 % .Dari keempat metode tersebut terdapat 2 metode dengan akurasi tertinggi adalah metode Decision Tree(ID3) yaitu 99,05 % . Hal ini membuktikan bahwa metode yang tepat untuk prediksi kelulusan siswa adalah metode Decision Tree (ID3) .

Berikut merupakan table perbandingan AUC



Gambar 14. Grafik perbandingan AUC antara algoritma Naïve Bayes , K-NN , Decision Tree (ID3) , dan SVM.



Grafik diatas menunjukkan AUC (optimistic) tertinggi yaitu metode K-NN dengan nilai yang sama yaitu 1,000 , AUC tertinggi yaitu metode Naïve Bayes dengan nilai 0,996 , AUC (pessimistic) tertinggi yaitu Naïve Bayes dengan nilai 0,996.

PENUTUP

1. Kesimpulan

Pada penelitian ini dilakukan pemodelan menggunakan algoritma Naïve Bayes , K-NN , Decision Tree (ID3) , dan SVM dengan menggunakan data nilai siswa dari semester 1 sampai 5 dan Ujian Sekolah. Fokus penelitian ini adalah mencari algoritma terbaik dalam menentukan prediksi kelulusan siswa. Validasi model menggunakan X-Validation dan evaluasi model menggunakan confusion matrix . Dari perhitungan manual dan pengujian menggunakan RapidMiner menunjukkan hasil yang sama.

Hasil penelitian menunjukan bahwa model Decision Tree (ID3)memiliki akurasi tertinggi yaitu 99,05 % dibandingkan model Naïve Bayes dengan akurasi 81,90 % , K-NN 98,10% dan model SVM dengan akurasi 95,71% .

Dengan demikian terbukti bahwa untuk prediksi kelulusan siswa, maka algoritma yang terbaik adalah Decision Tree (ID3)dengan nilai akurasi yaitu 99,05 %.

b.Saran

Berdasarkan kesimpulan di atas maka saran-saran yang dapat diberikan dalam penelitian ini yaitu:

1. Sebagai pendidik guru di harapkan dapat memperhatikan prestasi dan semangat belajar siswa dengan memberikan masukan dan motivasi yang baik , sehingga siswa akan merasa terus bersemangat dalam menjalani proses belajar.
2. Pihak sekolah memberikan arahan kepada orang tua siswa agar selalu mengawasi anaknya ketika di rumah dengan membagi waktu antara belajar dan bermain , selalu memberikan

semangat dan dorongan serta motivasi yang baik yang dapat meningkatkan prestasi siswa.

3. Meningkatkan keamanan data-data siswa dan melakukan antisipasi segala jenis kemungkinan dengan melakukan pencegahan dengan cara membackup data-data siswa.

DAFTAR PUSTAKA

- Jaya Sumpena.,Nina Kurnia H.(2019), Analisis Prediksi Kelulusan Siswa PKBM paket C dengan metode algoritma Naïve Bayes. Jurnal TEDC Vol. 13 No. 2, Mei 2019
- Moh.Zainuddin.(2019), Perbandingan 4 algoritma berbasis particle swarm optimization (PSO) untuk prediksi kelulusan tepat waktu mahasiswa. Jurnal Vol.13, No. 1, Tahun 2019
- Putu Gede Surya Cipta Nugraha, I Wayan Aribawa, I Putu Okta Priyana, Gede Indrawan.(2016), penerapan metode decision tree (data mining) untuk memprediksi tingkat kelulusan siswa SMPN 1 KINTAMANI. Jurnal Seminar Nasional Vokasi dan Teknologi (SEMNASVOKTEK). Denpasar-Bali, 22 Oktober 2016
- Aries Saifudin.(2017), metode data mining untuk seleksi calon mahasiswa pada penerimaan mahasiswa baru di universitas pamulang. Jurnal: .umj.ac.id/index.php/jurtek Volume 10 No. 1 Januari 2018
- Rizqy Agung Permana.(2016), seleksi atribut pada metode support vector machine untuk menentukan kelulusan mahasiswa e-learning. Jurnal Evolusi Volume 4 Nomor 1 - 2016 - lppm3.bsi.ac.id/jurnal
- Suhardjono, Ganda Wijaya, Abdul Hamid (2019) , prediksi waktu kelulusan mahasiswa menggunakan SVM berbasis PSO. Jurnal Vol. 7 No. 2 – Tahun 2019
- Arif Pratama, Randy CahyaWihandika, Dian Eka Ratnawati (2018) Implementasi Algoritme Support

Vector Machine(SVM) untuk
Prediksi Ketepatan Waktu Kelulusan
Mahasiswa. Jurnal Vol. 2, No. 4,
April 2018