

Prediction of Student Graduation Time Using The Best Algorithm

¹Verry Riyanto, ²Abdul Hamid, ³Ridwansyah

^{1,2}Departement of Informatic Technology, STMIK Nusa Mandiri Jakarta

³Departement of Management Informatics, AMIK BSI Bogor

^{1,2}Departement of Informatic Technology, STMIK Nusa Mandiri Jakarta

Email: ¹ferry@gmail.com, ²hamid@gmail.com, ³rdwansyah@gmail.com

Article Info

Article history:

Received Nov 12th, 2018

Revised Jan 15th, 2019

Accepted Mar 02th, 2019

Keyword:

Decision Tree

Neural Network

Student Graduation

Support Vector Machine

The Best Algorithm

ABSTRACT

Data mining has a very important role in the world of education that can help educational institutions in predicting and making decisions related to student's academic status. In predicting student success in the future, a model that can predict well is needed. Decision Tree Model (DT) is a decision model that can predict clearly but has the disadvantage of not being able to accommodate large data, neural network (NN) is a very popular model because it uses non-linear data and can hold large data while support vector machine (SVM) can generalize from nonlinear to linear. Of the three methods, each has weaknesses and strengths, we use the NN, SVM, and DT algorithms to predict the graduation time of academic students in one of the private universities in Indonesia. The results of this study indicate that the three models produce accuracy of more than 80%, the DT model has an accuracy of 84.96%, NN has an accuracy of 84.68% and the SVM model has an accuracy of 85.18% higher than the other two models. The results arising from this study provide important reference material for planning the future success of students and faculty in early warning to students in the future.

Copyright © 2019 Puzzle Research Data Technology

Corresponding Author:

Ridwansyah,

Departement of Informatic Technology,

STMIK Nusa Mandiri Jakarta,

Jl. Damai No.8 Warung Jati Barat Margasatwa, Jakarta Selatan

Email: rdwansyah@gmail.com

DOI: <http://dx.doi.org/10.24014/ijaidm.v2i1.6424>

1. PENDAHULUAN

Universitas atau akademik yang menjadi produk utamanya adalah mahasiswa. Setelah lulus para mahasiswa dapat meneruskan ke jenjang berikutnya atau menjadi tenaga kerja untuk industri, usaha pemerintahan atau usaha sektor swasta, bahkan dapat pula menjadi seorang wiraswasta. meningkatnya permintaan terhadap pekerja terampil saat ini terutama dari mahasiswa, pendidikan menjadi batu loncatan dalam mengamankan pekerjaan dengan perspektif jangka panjang. Maka memprediksi prestasi akademik siswa sangat penting bagi institusi pendidikan karena program tersebut strategis dan dapat direncanakan untuk memperbaiki atau mempertahankan kinerja siswa dengan prestasi akademik selama masa studi di institusi. Sebagian besar lembaga pendidikan saat ini secara digital mengelola data siswa mereka yang dikelola pada server yang berada di kampus. Sistem yang menyimpan informasi siswa dalam bentuk nilai numerik dan hanya menyimpan dan mengambil informasi apa yang dikandungnya merupakan sistem yang tidak memiliki kecerdasan untuk menganalisa data.

Pada dasarnya kelulusan mahasiswa dengan tepat waktu dapat dipengaruhi dari berbagai jenis atribut yaitu jenis kelamin, jurusan SLTA, asal SLTA dan IPK dari semester awal sampai semester akhir dikarenakan atribut-atribut tersebut sangatlah penting didalam memprediksi keberhasilan kelulusan mahasiswa. Untuk memberikan suatu keputusan prediksi kelulusan mahasiswa dapat dilakukan secara tepat waktu atau tidak dengan menyajikan sebuah pendekatan serta mengklasifikasikan mahasiswa untuk mencapai nilai akhir mereka. Kinerja akademik dalam penelitian ini diukur dengan nilai rata-rata indeks prestasi kumulatif dari semester awal sampai akhir atau sampai lulus. Kami melakukan studi empiris menggunakan data lulusan

akademik saat ini dengan memprediksi mahasiswa dari tahun pertama untuk mencegah hasil nilai yang buruk seseorang mahasiswa. Proses data mining yang tersembunyi pada data masa lalu siswa dan digunakan untuk prediksi kelulusan siswa sebagai bagian dari proses pengambilan keputusan. Model untuk prediksi yang banyak dilakukan oleh para peneliti diantaranya Naïve Bayes (NB), Neural Network (NN), Decision Tree (DT) dan Support Vector Machine (SVM). Kami menguji tiga model, yaitu NN, DT dan SVM dalam memprediksi mahasiswa pada saat hasil kelulusan mereka.

Decision Tree merupakan salah satu tipikal teknik pemrosesan untuk memprediksi beberapa sub kelompok dari sebuah grup objek dengan mengamati relasi dan berbentuk pohon keputusan khusus untuk sebuah model keputusan, oleh karena itu DT memiliki keuntungan untuk dipahami dan dijelaskan aliran analisis dengan mudah tetapi memiliki kelemahan dengan tidak dapat menampung data yang cukup besar [1]. Neural Network atau disebut juga NN merupakan metode untuk memberikan prediksi yang akurat dengan tidak mempunyai batasan apriori [2], dan memiliki prediksi nonlinear yang memiliki performa terbaik dimana basis nonlinear berfungsi untuk mengekskspikan kemungkinan interaksi yang kuat antar variabel tetapi memiliki kelemahan dengan menggunakan data yang besar untuk pelatihan [3][4]. Support Vector Machine atau disebut juga SVM adalah metode yang akurat karena dapat digeneralisasikan ke non-linear [5], dan memiliki kemampuan belajar dan generalisasi yang lebih baik [6].

Dari kelemahan dan kelebihan ketiga metode diatas maka dapat menjadi sebuah masalah dalam mencari arsitektur yang cocok dan bobot jaringan yang sesuai untuk sebuah model keputusan dalam memprediksi kelulusan mahasiswa yang sangat kompleks, dengan demikian tujuan penelitian ini adalah untuk mencari arsitektur yang cocok dan pengetahuan yang signifikan untuk membantu dan mencari tahu faktor-faktor terpenting dalam keberhasilan antara beberapa model yang diujikan dengan sebuah data kelulusan mahasiswa yang sangat kompleks yang dapat meningkatkan sistem peringatan dini dengan memperkirakan keberhasilan akademik mahasiswa dimasa yang akan datang.

2. BAHAN DAN METODE

Peneliti di USA dengan menganalisis faktor yang mempengaruhi keberhasilan mahasiswa berdasarkan indeks prestasi kumulatif (IPK) diketahui bahwa hubungan nilai IPK tidak signifikan[7]. Data mining dalam memprediksi mahasiswa dengan menggunakan ketiga metode NB, DT dan NN, dari hasil ketiga metode tersebut maka diperoleh NN dapat memprediksi lebih baik dibandingkan dengan metode lainnya[8]. Prediksi kinerja siswa dengan membandingkan metode seperti regresi, DT, NN dan NB telah dilakukan dan mempunyai hasil bahwa metode NB memberikan akurasi yang lebih tinggi dibandingkan metode lainnya[9].

Penelitian yang kami lakukan adalah penelitian eksperimen dengan mengkomparasi dan mengevaluasi dari ketiga model NN, DT dan SVM untuk mengetahui ketiga model manakah yang memiliki akurasi yang paling tinggi. Data yang dikumpulkan dan di uji oleh kami adalah data primer dengan data kelulusan mahasiswa dari salah satu universitas yang ada di Indonesia, dengan data tersebut kami melakukan pengolahan data awal yang berupa sumber data yang di random dan pembentukan variable seperti jenis kelamin, jurusan slta, asal slta, ipk semester 1, ipk semester 2, ipk semester 3, ipk semester 3, ipk semester 4, ipk semester 5, dan ipk semester 6. Data atribut yang diperoleh sebanyak 9 atribut prediktor dan 1 atribut tujuan yaitu lulus tepat waktu atau tidak, dari data tersebut dilakukan uji dan eksperimen dengan metode dan model yang di usulkan. Setelah diuji dan di eksperimen dari ketiga model tersebut akan dikomparasi tingkat akurasi untuk mengetahui model manakah yang paling tinggi akurasi, dari hasil yang di uji maka terpilihlah model yang memiliki akurasi paling tinggi.

3. HASIL DAN PEMBAHASAN

Kami dalam penelitian menggunakan aplikasi rapid miner studio dengan menggunakan spesifikasi komputer CPU Intel(R) Core(TM) i5-4210U dengan RAM 4GB dan sistem operasi microsoft windows 7 32-bit. Dan dengan data Kebutuhan dasar dari masalah prediksi waktu kelulusan adalah adanya data sebelumnya atau data masa lalu berdasarkan masa depan yang akan diprediksi yaitu dengan data yang dapat dilihat pada Tabel 1.

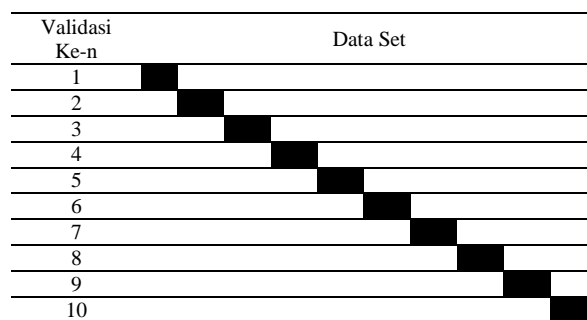
Tabel 1. Data Mahasiswa Kelulusan

JK	JRS_SLTA	ASAL_SLTA	IPK1	IPK2	IPK3	IPK4	IPK5	IPK6	TEPAT WAKTU
L	IPA	SMA N I LAIS	2.55	2.61	2.64	2.74	2.97	3.00	Ya
L	Teknik Komputer Jaringan	SMK TI YPML	3.23	3.15	3.07	3.17	3.25	3.17	Ya
P	Administrasi Perkantoran	SMK Nusantara 1	2.73	2.73	2.62	2.64	2.89	3.00	Ya
L	IPA	SMA YUPPENTEK 1	2.73	3.1	2.97	3.16	3.3	3.33	Ya
L	IPS	Hidayatul Ikhwan	2.18	2.44	2.11	2.36	2.66	2.57	Tidak
P	TKJ	SMK PUSTEK	2.32	2.39	2.43	2.74	2.67	2.95	Tidak
P	Penjualan	SMK Negeri 1	3.64	3.56	3.59	3.53	3.53	3.51	Ya

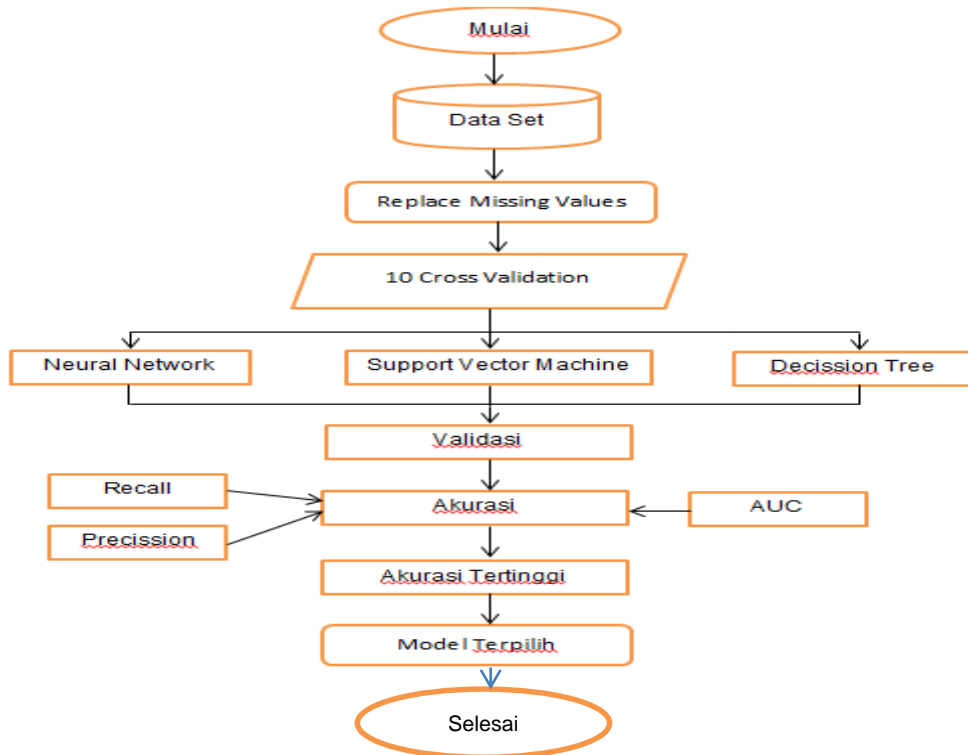
JK	JRS_SLTA	ASAL_SLTA	IPK1	IPK2	IPK3	IPK4	IPK5	IPK6	TEPAT WAKTU
L	Otomotif	SMK Negeri 2	2.5	2.71	2.57	2.58	2.75	2.94	Tidak
P	IPA	SMA Negeri 1 Jiwan	2.91	3.07	2.84	2.95	3.16	2.98	Tidak
L	Teknik Komputer	SMKN 1 Panongan	3.09	2.88	2.89	3.02	3.1	3.15	Ya
L	Multimedia	SMK Pancakarya	2.86	2.88	3.10	3.11	3.14	3.16	Ya
L	Teknik Listrik Industri	SMKN 7 (STM Pembangunan)	3.64	3.71	3.80	3.83	3.77	3.77	Ya
L	IPS	SMU Negeri 6	2.91	3.05	2.74	2.70	2.65	2.69	Tidak
P	Ilmu Pengetahuan Sosial	Madrasah Aliyah	2.36	2.17	2.08	2.46	2.89	2.91	Ya
L	IPA	SMA Paradigma	2.73	2.83	2.9	2.88	2.88	2.91	Ya
L	Teknik Komputer Jaringan	SMKN 1 Panongan Kab.Tange	3.45	3.37	3.46	3.49	3.49	3.56	Ya
P	IPA	SMA Methodist	2.68	2.80	2.95	3.09	3.13	3.14	Ya
P	IPA	SMA Negeri 2 Binjai	3.32	3.49	3.56	3.56	3.63	3.61	Ya
P	IPA	Al-Husna	3.09	3.12	2.98	3.05	3.35	3.40	Ya
L	IPA	SMA Swasta Yaspih Rajeg	3.41	3.29	3.13	3.10	3.10	3.18	Ya
L	IPA	SMA Negeri 1 Lemahabang	2.73	2.78	2.92	3.06	3.12	3.13	Ya
L	Teknik Listrik	SMKN 1 Kab.Tangerang	3.09	3.12	3.05	3.28	3.32	3.37	Ya
L	Otomotif	SMK Pancasila 3 Baturetno	3.09	3.00	2.95	3.07	3.08	3.14	Ya
P	Ilmu Pengetahuan Alam	SMA Negeri 1 Karangdowo	3.73	3.66	3.66	3.69	3.66	3.57	Ya
L	Ilmu Pengetahuan Alam	SMA Negeri 11 Tangerang	3.23	3.34	3.03	3.23	3.19	3.22	Ya
L	Animasi	Bina Informatika	2.86	2.85	3.07	3.11	3.13	3.09	Ya

Model algoritma yang diusulkan dapat dilihat gambar 1. Dengan proses perhitungan untuk memprediksi kelulusan mahasiswa dapat diuraikan dengan data kelulusan mahasiswa dari total 796 mahasiswa pada salah satu akademik swasta yang ada di Indonesia digunakan dalam penelitian ini dengan metode DT, NN dan SVM adalah data primer yang kita uji. Data siswa-siswa ini berasal dari data mahasiswa berbagai kalangan, data terdiri dari 796 parameter atribut diukur dan di prediksi dengan sembilan atribut inputan dan satu atribut tujuan. Atribut input yang kami gunakan diantaranya jenis kelamin (JK), jurusan SLTA (JRS_SLTA), asal slta (ASAL_SLTA), IPK semester satu (IPK1), IPK semester dua (IPK2), IPK semester tiga (IPK3), IPK semester empat (IPK4), IPK semester lima (IPK5), IPK semester enam (IPK6), dan atribut tujuan yaitu tepat waktu. Dari data primer tersebut akan dilakukan pengecekan data nilai yang hilang, setelah pengecekan data nilai yang hilang dilakukan model parameter yang dioptimalkan berdasarkan 10 cross validation yang dapat dilihat pada tabel 2.

Tabel 2. Ilustrasi 10 Fold Cross Validation



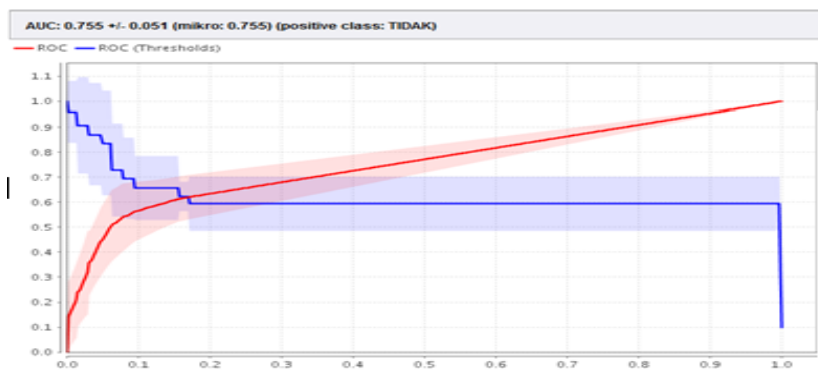
Setelah dilakukan proses 10 cross validation maka dilakukan uji dengan ketiga model yaitu DT, NN dan SVM. Dari hasil uji coba dengan ketiga model maka didapatkan hasil yang diperoleh dibandingkan dalam hal validasi dengan akurasi, recall, precision dan auc. Dari hasil tersebut maka akurasi yang tertinggi lah yang menjadi model terpilih.



Gambar 1. Flow Diagram Metode Usulan

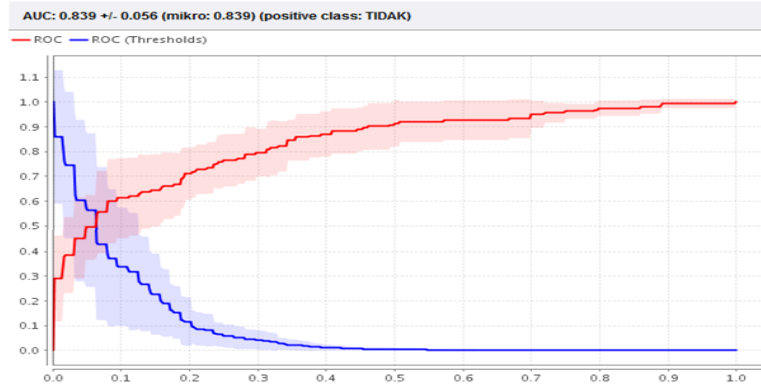
Dari uji coba yang dilakukan pada model algoritma decision tree maka didapatkan akurasi sebesar 86.19%, precision 74.80%, recall 55.51% dan AUC sebesar 0.755%. Maka didapatkan hasil perhitungan akurasi, kurva ROC dan arsitektur decision tree pada Gambar 2 dan Gambar 3.

$$\begin{aligned}
 \text{Akurasi} &= \frac{(TN + TP)}{(TN + FN + TP + FP)} \dots\dots\dots(1) \\
 &= \frac{(596 + 90)}{(596 + 38 + 90 + 70)} \\
 &= 0,8619
 \end{aligned}$$

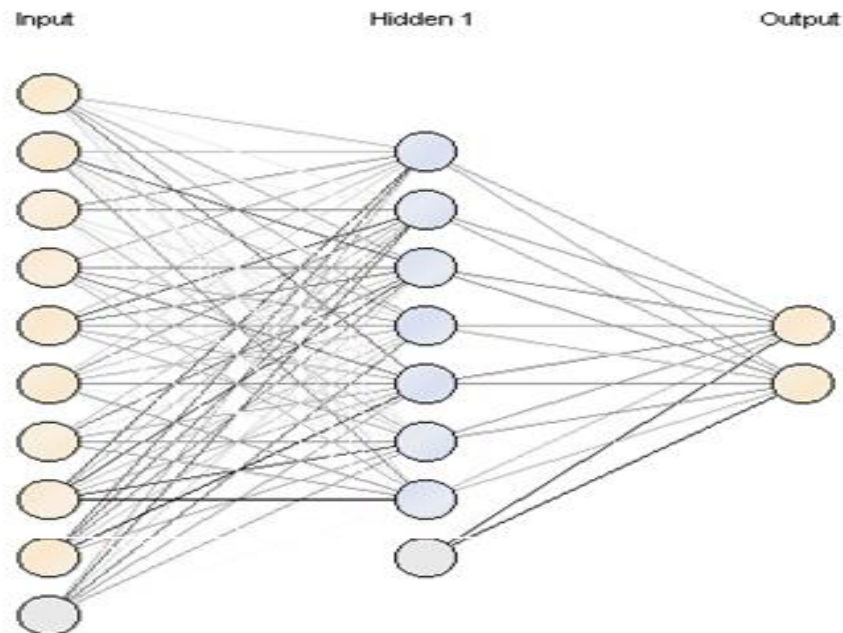


Gambar 2. Kurva ROC Decision Tree

$$\begin{aligned}
 \text{Akurasi} &= \frac{(TN + TP)}{(TN + FN + TP + FP)} \dots\dots\dots(2) \\
 &= \frac{(596 + 78)}{(596 + 38 + 78 + 84)} \\
 &= 0,8468
 \end{aligned}$$



Gambar 4. Kurva ROC Neural Network



Gambar 5. Arsitektur Neural Network

Pada arsitektur neural network yang sudah dilakukan didapatkan nilai bobot *hidden layer (sigmoid)* dapat dilihat pada Tabel 3 dan output (sigmoid) dapat dilihat pada Tabel 4.

Tabel 3. Nilai Bobot Hidden Layer (Sigmoid)

Node	Output (Sigmoid)						
	1	2	3	4	5	6	7
JK	4.303	1.236	5.515	-5.935	-9.748	1.492	2.459
JRS SLTA	4.340	-1.023	10.265	-5.915	1.423	-1.722	-7.277
ASAL SLTA	6.656	-7.048	-0.201	1.988	1.368	-5.126	-1.419
IPK1	-5.685	-0.708	-7.002	-5.618	8.757	2.854	5.175
IPK2	-4.090	-11.289	8.344	-5.898	5.026	-5.116	-0.341
IPK3	-2.926	-7.599	3.276	4.732	-7.708	-0.961	6.686
IPK4	-0.206	4.688	9.701	1.514	5.330	6.551	-6.090
IPK5	11.450	4.800	13.263	3.890	4.805	11.666	15.932
IPK6	9.544	-9.591	-8.065	-3.710	-15.018	5.651	0.174
Bias	-4.708	-10.112	2.711	-9.957	-6.843	-6.460	-0.254

Dari tabel 3 ditunjukkan bahwa terdapat delapan kolom dimana kolom pertama merupakan atribut pada input layer dan kolom lainnya merupakan simpul *hidden layer*.

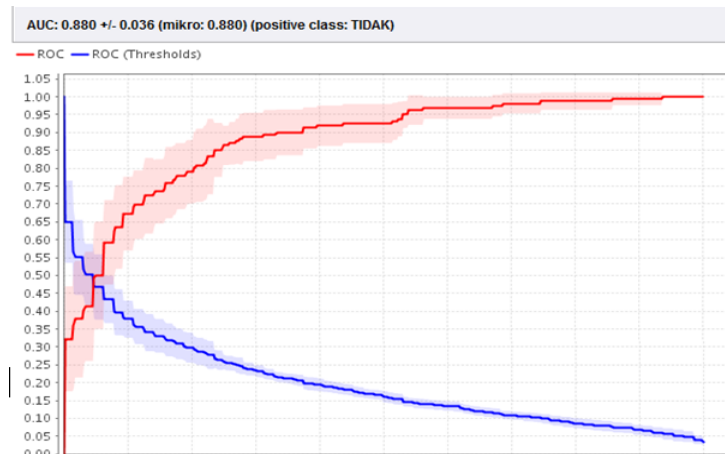
Tabel 4. Nilai Bobot Akhir Output Layer (Sigmoid)

Output (Sigmoid)	Class Ya	Class Tidak
1	6.347	-6.347
2	6.401	-6.401
3	7.470	-7.470
4	5.280	-5.280
5	7.840	-7.840
6	6.537	-6.537
7	4.618	-4.618
Threshold	-15.346	15.346

Dari tabel 4 menunjukkan bahwa terdapat tiga kolom dimana kolom kedua dan ketiga merupakan *class* atau atribut class dan kolom pertama merupakan nilai bias.

Selanjutnya kami melakukan uji coba menggunakan algoritma support vector machine dan mendapatkan hasil akurasi sebesar 85.18%, precision 79.58%, recall 40.22% dan AUC sebesar 0.880%. Kami mencari hasil perhitungan akurasi menggunakan rumus yang digunakan model support vector machine dan mendapatkan kurva roc pada gambar 6.

$$\begin{aligned}
 \text{Akurasi} &= \frac{(TN + TP)}{(TN + FN + TP + FP)} \dots\dots\dots(3) \\
 &= \frac{(613 + 65)}{(613 + 21 + 65 + 97)} \\
 &= 0,8518
 \end{aligned}$$



Gambar 6. Kurva ROC Support Vector Machine

Dengan melakukan uji coba pada SVM maka didapatkan nilai dari kernel model SVM

- Bias (offset): -1.318
- w[JK] = -0.070
- w[JRS_SLTA] = -0.019
- w[ASAL_SLTA] = 0.099
- w[IPK1] = -0.010
- w[IPK2] = 0.045
- w[IPK3] = 0.061
- w[IPK4] = -0.316
- w[IPK5] = -1.135
- w[IPK6] = 0.250

Dari ketiga model yang telah diuji maka memiliki perbandingan dari ketiga model tersebut yang dapat dilihat pada Tabel 5.

Tabel 5. Nilai Perbandingan

Nilai/ Hitung	DT	NN	SVM
Akurasi	84.96%	84.68%	85.18%
Precision	66.67%	71.13%	79.58%
Recall	60.00%	48.27%	40.22%
AUC	0.831%	0.839%	0.880%

Dari Tabel 5 yang merupakan hasil komparasi dengan membandingkan akurasi, precision, recall dan AUC. Maka nilai akurasi untuk model svm memiliki tingkat akurasi tertinggi diantara model lainya dengan nilai 85.18%, begitu pula dengan precisionnya dengan nilai 79.58%, dan untuk nilai AUC tertinggi dengan nilai 0.880%. Namun untuk nilai recall dan kurva roc svm paling rendah dibandingkan dengan yang lain. Untuk nilai recall tertinggi 60.00% yang dimiliki oleh DT. Maka dengan menguji dari ketiga model tersebut, metode svm yang cocok untuk menguji data kelulusan mahasiswa.

4. KESIMPULAN

Waktu kelulusan mahasiswa sangat berpengaruh terhadap sebuah institusi di bidang pendidikan, dengan hasil yang didapat dari penelitian prestasi akademik selama masa studi di institusi, menggunakan arsitektur dan bobot jaringan yang cocok dari ketiga metode NN, DT dan SVM. Dari ketiga model di ujikan dan di eksperimen dengan nilai akurasi tertinggi yaitu SVM 85,18% untuk posisi kedua dengan nilai 84,96% di tempati oleh DT, dan nilai akurasi paling kecil yaitu NN dengan nilai 84,68%. maka model yang paling cocok untuk memprediksi waktu kelulusan mahasiswa dengan akurasi paling tinggi yaitu model SVM dengan hasil nilai 85,18%, maka sistem dengan menggunakan model sekarang dapat memiliki kecerdasan untuk menganalisa data, untuk mencegah mahasiswa mendapatkan hasil nilai yang buruk sehingga dapat lulus tepat waktu dan universitas mendapatkan pengetahuan yang signifikan untuk meningkatkan sistem peringatan dini bagi mahasiswanya.

REFERENCES

- [1] J. K. Choi, C. Il Yoo, K. A. Kim, Y. Won, and J. J. Kim, "Study on datamining technique for foot disease prediction," in *International Conference on IT Convergence and Security, ICITCS 2014*, 2014.
- [2] S. Moro and R. M. S. Laureano, "Using Data Mining for Bank Direct Marketing: An application of the CRISP-DM methodology," *Eur. Simul. Model. Conf.*, pp. 117–121, 2011.
- [3] S. Moro, P. Cortez, and P. Rita, "A data-driven approach to predict the success of bank telemarketing," *Decis. Support Syst.*, vol. 62, pp. 22–31, 2014.
- [4] Ridwansyah and E. Purwaningsih, "Particle Swarm Optimization Untuk Meningkatkan Akurasi Prediksi Pemasaran Bank," *J. PILAR Nusa Mandiri*, vol. 14, no. 1, pp. 83–88, 2018.
- [5] M. Sugimoto, M. Takada, and M. Toi, "Comparison of robustness against missing values of alternative decision tree and multiple logistic regression for predicting clinical data in primary breast cancer," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, 2013, pp. 3054–3057.
- [6] J. L. Tang, Q. R. Cai, and Y. J. Liu, "Prediction of material mechanical properties with Support Vector Machine," in *International Conference on Machine Vision and Human-Machine Interface*, 2010, vol. 1, pp. 592–595.
- [7] Freitas F. A. and L. L. J., "Maslow's hierarchy of needs and student academic success," *Teach. Learn. Nurs.*, vol. 6, no. 1, pp. 9–13, 2011.
- [8] M. V. Ashok and A. Apoorva, "Data mining approach for predicting student and institution's placement percentage," in *2016 International Conference on Computation System and Information Technology for Sustainable Solutions, CSITSS 2016*, 2016, pp. 336–340.
- [9] M. T. Devasia, M. V. T. P, and M. V. Hegde, "Prediction of Students Performance using Educational Data Mining," *Int. J. Cogn. Ther.*, vol. 1, no. 3, pp. 266–279, 2008.

BIBLIOGRAFI PENULIS

Verry Riyanto, Gelar Diploma Tiga didapat dari AMIK BSI Bekasi Jurusan Manajemen Informatika sejak tahun 2009, Kemudian Sarjana Komputer diperoleh dari Program Studi Sistem Informasi pada STMIK Nusa Mandiri Jakarta tahun 2010, lalu Gelar Magister Ilmu Komputer diperoleh dari Program Studi Ilmu Komputer pada Pascasarjana STMIK Nusa Mandiri Jakarta tahun 2014. Saat ini aktif mengajar sebagai Dosen pada Sekolah Tinggi STMIK Nusa Mandiri Jakarta dan Bina Sarana Informatika.



Abdul Hamid sebagai tenaga pengajar pada AMIK BSI Bogor, Prodi Manajemen Informatika



Ridwansyah sebagai tenaga pengajar pada STMIK Nusa Mandiri Jakarta Prodi Teknik Informatika. Memperoleh gelar Sarjana Komputer pada tahun 2011 di STMIK Nusa Mandiri Jakarta, prodi Sistem Informasi. Memperoleh gelar Magister Komputer pada STMIK Nusa Mandiri Jakarta tahun 2014. Pernah melakukan penelitian dengan tema kecerdasan buatan dan data mining.