

PREDIKSI KELULUSAN TEPAT WAKTU MENGGUNAKAN ALGORITMA SVM DAN K-NEAREST NEIGHBOUR BERBASIS PARTICLE SWARM OPTIMIZATION DI STMIK ERESHA

Imam Hidayat¹

¹Universitas Pamulang, Jalan Surya Kencana No.1, Pamulang Barat, Pamulang, Kota Tangerang Selatan, Banten 15417, Indonesia
Program Studi Sistem Informasi Fakultas Teknik Universitas Pamulang

*E-mail: imam.males99@gmail.com

ABSTRAK

PREDIKSI KELULUSAN TEPAT WAKTU MENGGUNAKAN ALGORITMA SVM DAN K-NEAREST NEIGHBOUR BERBASIS PARTICLE SWARM OPTIMIZATION DI STMIK ERESHA. Kelulusan mahasiswa bagi suatu perguruan tinggi menjadi elemen penting untuk penilaian akreditasi. Kelulusan tepat waktu untuk pencapaian gelar magister membutuhkan waktu normal selama 2 tahun atau 4 semester, namun pada kenyataannya banyak mahasiswa yang lulus tidak tepat waktu. Banyak faktor yang menyebabkan ketidaktepatan waktu kelulusan antara lain bersumber dari faktor *internal* dan faktor *external*. Pada penelitian ini telah diterapkan metode klasifikasi dengan algoritma *Support Vector Machine* dan *K-Nearest Neighbour* berbasis *Particle Swarm Optimization* untuk mencari metode yang terbaik dalam mengklasifikasi data mahasiswa di STMIK Eresha. Terkait atribut yang digunakan adalah IPS (Indek Prestasi Sementara), SKS (Satuan Kredit Semester) lulus, usia, asal perguruan tinggi, status mahasiswa, status pernikahan dan status pekerjaan mahasiswa. Hasil penelitian ini menunjukkan bahwa algoritma terbaik antara algoritma SVM dan K-NN yaitu algoritma K-NN dengan nilai k-7 dengan nilai akurasi 78,73%, sedangkan algoritma SVM sebesar 73,41% dan setelah penambahan algoritma PSO akurasi terbaik terdapat pada nilai K-9 dengan akurasi 80,21% dengan penambahan nilai akurasi sebesar 1,48 %. Hasil penelitian ini menunjukkan bahwa algoritma K-NN menjadi algoritma yang tepat dibandingkan dengan algoritma SVM untuk memprediksi kelulusan tepat waktu di STMIK Eresha.

Kata Kunci: Kelulusan tepat waktu, Prediksi, SVM, K-NN, PSO

ABSTRACT

Student graduation for university becomes an important element for accreditation assessment. Timely graduation to achieve a master's degree requires normal time for 2 years or 4 semesters, but in reality many students do not graduate on time. Many factors that cause the inaccuracy of graduation time include internal and external factors. In this study a classification method with the Support Vector Machine and K-Nearest Neighbour s based on Particle Swarm Optimization has been applied to find the best method for classifying student data at Eresha STMIK. Related to the attributes used are GPA (Great Point Average), credit passed, age, university origin, student status, marital status and student employment status. The results of this study indicate that the best algorithm between the SVM and K-NN algorithm is the K-NN algorithm with k-7 value with an accuracy value of 78.73%, while the SVM algorithm is 73.41% and after the addition of the PSO algorithm the best accuracy is found in the value K-9 with an accuracy of 80.21% with the addition of an accuracy value of 1.48%. The results of this study indicate that the K-NN algorithm is the right algorithm compared to the SVM algorithm to predict timely graduation at STMIK Eresha.

Keywords: Timely graduation prediction, SVM, -NN, PSO

I. PENDAHULUAN

Perguruan tinggi memiliki peranan penting dalam menyumbang sumber daya manusia (SDM) disuatu negara oleh karena itu perguruan tinggi dituntut untuk bersaing dengan sumber daya yang dimiliki, selain sumber daya, sarana, prasarana, dan manusia sistem informasi juga mempunyai daya saing yang tinggi dalam suatu perguruan tinggi. Kelulusan mahasiswa bagi suatu perguruan tinggi menjadi elemen penting untuk penilaian akreditasi (Yalidhan, 2018). Bisa dikatakan bahwa kelulusan tepat waktu mahasiswa menjadi tolok ukur suatu perguruan tinggi dikatakan sukses atau tidaknya.

Kelulusan tepat waktu adalah salah satu syarat

mutlak bagi mahasiswa magister komputer yang mendapat beasiswa (ikatan dinas), dimana para lulusan magister komputer ini akan direkomendasikan menjadi dosen di STMIK Eresha atau UNPAM (Universitas Pamulang). Pencapaian gelar magister tersebut membutuhkan waktu normal selama 2 tahun atau 4 semester, akan tetapi pada kenyataannya banyak mahasiswa magister komputer ini tidak dapat menyelesaikan studinya tepat pada waktu yang telah ditentukan. Banyak faktor yang menyebabkan ketidaktepatan waktu kelulusan tersebut, faktor-faktor tersebut dapat bersumber dari faktor *internal* dan faktor *external* (Zainuddin, 2019). Hal ini menjadi menarik untuk dijadikan penelitian. Apakah faktor *internal* mahasiswa atau *external* (kampus) yang menyebabkan mahasiswa magister computer tidak lulus tepat waktu.

Belum adanya metode yang tepat untuk

melakukan prediksi kelulusan tepat waktu di STMIK Eresha bagi mahasiswa magister komputer, baik mahasiswa reguler maupun yang mendapatkan beasiswa, mendasari untuk dilakukannya penelitian ini, walaupun begitu penelitian tentang prediksi kelulusan tepat waktu bukan suatu hal yang baru karena sudah banyak yang melakukan penelitian sejenis ini untuk objek yang berbeda. Penelitian seperti ini pernah dilakukan untuk memprediksi kelulusan mahasiswa magister komputer Universitas Amikom Yogyakarta (Susanto & Fatta, 2018). Metode yang digunakan dalam penelitian ini adalah K-NN.

Objek yang diteliti adalah data mahasiswa tahun 2014 sampai dengan 2015. Sedangkan atribut yang digunakan adalah nilai IPK semester 1-4, capaian SKS dan status kelulusan. *Output* dari sistem ini adalah kelulusan mahasiswa yang dibagi menjadi dua yaitu kelulusan tepat waktu dan kelulusan tidak tepat waktu. Hasil dari penelitian ini berdasarkan penerapan $k=14$ dan $k\text{-fold} = 5$ adalah performa terbaik dalam memprediksi kelulusan mahasiswa dengan metode K-NN menggunakan indeks prestasi

4 semester dengan nilai akurasi = 98,46 %
precision =

99.53 % dan recall = 97.64%. Hasil ini berbeda dengan penelitian yang dilakukan oleh Moh. Zainudin (Zainuddin, 2019) yang membandingkan 4 algoritma klasifikasi yang berbasis PSO (*particle swarm optimization*) untuk prediksi kelulusan tepat waktu, dengan perbedaan jumlah atribut yang lebih banyak dan hasilnya hanya mendapatkan akurasi sebesar 69.81 % pada metode K-NN namun setelah ditambahkan fitur PSO maka akurasinya meningkat

5.21 % atau menjadi 74.02 %. Dari perbedaan inilah

diharapkan ada metode yang tepat untuk memprediksi kelulusan tepat waktu di magister komputer SMTIK Eresha.

2. LANDASAN TEORI

2.1 . Penelitian Terkait

Penelitian yang dilakukan oleh (Susanto & Fatta,

2018) menyatakan bahwa memprediksi kelulusan mahasiswa dengan metode *K-Nearest Neighbour* menggunakan indeks prestasi 4 semester dengan nilai akurasi= 98,46%,
precision= 99.53% dan recall

=97.64%. Input dari sistem ini adalah data sampel berupa data mahasiswa tahun 2014-2015. Analisa metode ini menggunakan dua pengujian yaitu data *testing* dan data *training*. Kriteria yang digunakan dalam penelitian ini adalah IP Semester 1-4, capaian SKS dan Status Kelulusan. *Output* dari sistem ini berupa hasil prediksi kelulusan mahasiswa yang terbagi menjadi dua yaitu tepat waktu dan kelulusan tidak tepat waktu. Hasil pengujian berdasarkan penerapan $k = 14$ dan $k\text{-fold} = 15$ menunjukkan bahwa performa yang terbaik dalam memprediksi kelulusan mahasiswa

dengan metode *K-Nearest Neighbour* menggunakan indeks prestasi 4 semester dengan nilai akurasi= 98,46%, precision= 99.53% dan recall =97.64%.

Pada penelitian yang dilakukan oleh Zainudin dia membandingkan 4 algoritma yaitu algoritma *Naive Bayes*, *Decision Tree (C4.5)*, *K-Nearest Neighbour (k-NN)*, *Neural Network* berbasis *Particle Swarm Optimization (PSO)* sebagai referensi untuk membuat kebijakan dan tindakan bidang akademik (BAAK) dalam mengurangi mahasiswa yang lulus terlambat dan tidak lulus. Hasil menunjukkan Algoritma *K-Nearest Neighbour (k-NN)* berbasis *PSO* pada k-optimum=19 mempunyai performa terbaik dari 4 algoritma yang ada, dengan nilai *Accuracy* = 74,08% dan nilai *Area Under the Curve (AUC)* = 0,788. Penambahan fitur *Particle Swarm Optimization (PSO)* selalu meningkatkan nilai akurasi, dimana peningkatan nilai akurasi tertinggi terletak pada Algoritma *Decision Tree (C4.5)* sebesar 5,21%, terendah pada Algoritma *Naive Bayes* sebesar 2,13%. Dari hasil ini diketahui bahwa algoritma *Decision tree* meningkat nilai akurasinya setelah dilakukan optimasi dengan *PSO (Particle Swarm Optimization)*. Berkaitan untuk penelitian selanjutnya perlu dicoba dengan penambahan kuantitas (jumlah dataset) dan kualitas (jumlah atribut) yang akan digunakan untuk klasifikasi data mining serta dapat pula ditambahkan beberapa model klasifikasi yang baru misalnya *ID3*, *Support Vector Machine* untuk diuji coba untuk mengetahui tingkat *accuracy* dan *Area Under the Curve (AUC)* untuk klasifikasi bidang kerja lulusan mahasiswa.

Pada penelitian yang dilakukan oleh Arif Rakhman (Rakhman, 2017) dalam judulnya “Prediksi Ketepatan Kelulusan Mahasiswa Menggunakan Metode *Decision Tree* berbasis *Particle Swarm Optimization (PSO)*” data yang digunakan dalam penelitian ini antara lain jumlah data eksperimen 300 data dengan atribut seperti jenis kelamin, status pernikahan, status pekerjaan, *IPS_1*, *IPS_2*, *IPS_3*, *IPS_4*. Atribut-atribut tersebut diambil dari data sistem akademik mahasiswa Politeknik Harapan Bersama Tegal. Dari hasil analisis optimasi dapat disimpulkan bahwa nilai akurasi yang didapat pada model algoritma *decision tree* dengan *PSO* adalah 97.67 % lebih baik jika dibandingkan dengan model algoritma *decision tree* yaitu 96.67 %. Sehingga dapat disimpulkan bahwa penerapan teknik optimasi

Particle Swarm Optimization dapat meningkatkan nilai akurasi pada algoritma *decision tree*. Akurasi memang terlihat cukup besar akan tetapi ini hanya menggunakan data sebanyak 300 data. Apakah akan sama nilai akurasinya apabila data yang digunakan lebih banyak?

Ada banyak penelitian terkait prediksi ataupun terkait dengan metode klasifikasi, berikut saya jelaskan dalam bentuk tabel dibawah ini:

Tabel 2.1 Literatur terkait

1	Tahun	2017
	Judul	Metode Data Mining Untuk Seleksi Calon Mahasiswa Pada Penerimaan Mahasiswa Baru di Universitas Pamulang
	Penulis	Aries Saefudin
	Permasalahan	Sistem seleksi saat ini belum dapat mendeteksi kelulusan tepat waktu
	Metode	(NB), (k-NN), (RF), (DS), (DT), (RI), (LR), (LDA), (NN), dan (SVM).
	Hasil	Menghasilkan akurasi 65.00% pada algoritma SVM (<i>Support Vector Machine</i>) untuk prediksi ketepatan mahasiswa lulus tepat waktu
2	Tahun	2017
	Judul	Penerapan Teknik Klasifikasi Untuk Prediksi Kelulusan Mahasiswa Menggunakan Algoritma <i>K-Nearest Neighbour</i>
	Penulis	Yandi saputra dan Yogi Primadasa
	Permasalahan	Tidak stabilnya tingkat kelulusan mahasiswa program studi Sistem Informasi pada STMIK Bina Nusantara Jaya Lubuklinggau
	Metode	<i>K-Nearest Neighbour</i>
	Hasil	Hasil dari Index Prestasi semester kurang efektif dikarenakan hanya hasil dari Index Prestasi tidak dari nilai masing-masing mata kuliah

	Metode	SVM
3	Tahun	2018
	Judul	Implementasi Data Mining Menggunakan Model SVM Untuk Prediksi Kepuasan Pengunjung Taman Tabebuaya
		Agus darmawan dkk
	Permasalahan	Karena taman Tabebuaya lebih sepi dibanding taman lainnya, maka dibuat sistem prediksi kepuasan pelanggan.
	Metode	SVM
	Hasil	algoritma SVM memiliki akurasi dan performa secara rata-rata yaitu sebesar 86,00% dan nilai AUC sebesar 0.947
4	Tahun	2018
	Judul Prediksi Pergerakan Harga	Judul Prediksi Pergerakan Harga
	Penulis	Fiqi Ruli setiawan
	Permasalahan	Kesulitan Dalam Memprediksi Pergerakan Harga Saham
	Metode	SVM menggunakan Trend Deterministic Data
	Hasil	Tingkat akurasi menggunakan <i>trend deterministic data preparation</i> yaitu 52,06%, 55,52%, 52,06%

2.2 Prediksi

Prediksi adalah suatu usaha untuk meramalkan keadaan dimasa mendatang melalui pengujian keadaan dimasa lalu (linda 2018). Dalam kehidupan nyata segala sesuatu itu serba tidak pasti, kita hanya bisa memprediksi dari apa yang sudah terjadi dimasa lalu atau kejadian yang sudah terulang, walau demikian tidak mudah untuk memprediksi secara cepat segala kemungkinan yang akan terjadi. Dalam hal ini diperlukan *forecast*. *Forecasting* ini dibuat untuk meminimumkan masalah ketidakpastian pada suatu perusahaan, dengan artian bahwa *forecasting* akan menghasilkan *forecast* yang bisa meminimumkan kesalahan dalam meramal atau memprediksi yang biasanya diukur dengan *mean squad error*, *mean absolut error*, dan

sebagainya. Prediksi tidak harus memberikan jawaban secara pasti kejadian yang akan terjadi, melainkan berusaha untuk mencari jawaban sedekat mungkin dengan kejadian (Sinaga et al., 2018). Jadi penulis menyimpulkan bahwa prediksi adalah suatu ramalan tentang suatu kejadian yang akan terjadi dengan mempelajari hal hal yang berkaitan dengan kejadian tersebut dengan tujuan agar menghasilkan suatu ramalan yang mendekati benar. Prediksi dalam hal apapun memang dasarnya hanya menduga-duga, akan tetapi apabila kita sudah mempelajari hal-hal yang berkaitan dengan apa yang akan kita prediksi seperti hal yang sudah dilakukan orang lain, metode yang digunakan, kebiasaan atau sejarah maka hasil yang akan kita dapatkan akan maksimal atau mendekati kata benar.

2.3 Lulus Tepat Waktu

Lulus tepat waktu disini adalah kelulusan tepat waktu bagi mahasiswa magister komputer STMIK Eresha dalam menyelesaikan masa studinya atau masa kuliahnya, dikatakan lulus tepat waktu apabila mahasiswa dapat menyelesaikan dalam waktu 5 semester. Dalam upaya meningkatkan atau mempertahankan kinerja mahasiswa selama masa studi di perguruan tinggi, maka perlu diprediksi tingkat kelulusan mahasiswa di setiap tahunnya sebagai salah satu faktor yang menentukan kualitas perguruan tinggi tersebut, karena tingginya tingkat keberhasilan mahasiswa dan rendahnya tingkat kegagalan mahasiswa merupakan cermin kualitas dari suatu perguruan tinggi (maulana. devi, 2019)..

2.4. Data Mining

Secara sederhana data mining adalah penambangan atau penemuan informasi baru dengan mencari pola atau aturan tertentu dari sejumlah data yang sangat besar. Data mining adalah suatu istilah yang digunakan untuk menemukan pengetahuan yang tersembunyi di dalam database (Muhammad et al.,2018).

Istilah data mining memiliki beberapa padanan , seperti *knowledge discovery* ataupun *pattern recognition*. Kedua istilah tersebut sebenarnya memiliki ketepatannya masing-masing, istilah *knowledge discovery* atau penemuan pengetahuan tepat digunakan karena tujuan utama dari data mining memang untuk mendapatkan pengetahuan yang tersembunyi di dalam bongkahan data (Rakhman, 2017).

Dari beberapa kutipan tentang data mining diatas penulis dapat menyimpulkan bahwa data mining adalah pola untuk menemukan, mengolah dan mengoperasikan data dalam suatu *database* dalam jumlah yang besar.

2.5 Klasifikasi

Klasifikasi adalah pemrosesan untuk menemukan sebuah model atau fungsi yang menjelaskan dan mencirikan konsep atau kelas data, untuk kepentingan tertentu. Menurut Gorunescu, *Data Mining: Concepts, Models and Techniques (Intelligent Systems Reference Library)*. Yang ditulis dalam penelitian Abdur Rohman dkk menjelaskan bahwa proses klasifikasi didasarkan pada empat komponen: (Kelas, Predictor, Training dataset, Testing)

2.6 Support Vector Machine (SVM)

Support Vector Machine adalah metode *learning machine* yang bekerja atas prinsip *Structural Risk Minimization* (SRM) dengan tujuan menemukan *hyperlane* terbaik yang memisahkan dua buah *class* pada *input space*. SVM adalah salah satu algoritma klasifikasi yang dapat menghasilkan proses pembelajaran dalam suatu masalah klasifikasi di artikan sebagai upaya mencari garis (*hyperlane*) untuk memisahkan dari dua kelompok tersebut (Darmawan et al., 2018).

Inti dari metode ini adalah pencarian *hyperlane* terbaik dari setiap kemungkinan. Persamaan yang terletak pada SVM adalah sebagai berikut:

$$\bar{w}.x + b = -1 \quad w'.x + b = +1$$

metode klasifikasi. Terdapat dua kategori SVM yaitu *Support Vector Machine classification* dan *Support Vector Machine Regression*.

2.7 K-Nearest Neighbour (K-NN)

K-Nearest Neighbour (K-NN) adalah suatu metode yang menggunakan algoritma *supervised* dengan hasil dari *query instance* yang baru diklasifikasikan berdasarkan mayoritas dari label class pada K-NN (Muri et al., 2018)

Tujuan dari algoritma ini adalah untuk mengklasifikasikan obyek baru berdasarkan atribut dan *sample-sample* dari *training data*. Algoritma *K- Nearest Neighbour* menggunakan *Neighbourhood Classification* sebagai nilai prediksi dari nilai *instance* yang baru.

2.8 Particle Swarm Optimization (PSO)

Particle Swarm Optimization (PSO) adalah algoritma pencarian berbasis populasi dan diinisialisasi dengan populasi solusi acak yang disebut partikel. PSO merupakan metode pencarian yang berasal dari penelitian untuk gerakan sekelompok burung atau ikan. Serupa dengan algoritma genetik (GA), PSO melakukan pencarian menggunakan populasi (*swarm*) dari individu (partikel) yang akan diperbaharui dari iterasi ke iterasi (Rakhman, 2017).

3. ANALISIS KEBUTUHAN

3.1 Analisis Kebutuhan Data

Untuk kebutuhan data penelitian akan digunakan *dataset* mahasiswa magister komputer STMIK Eresha mulai dari tahun 2016 sampai tahun 2018. Adapun *dataset* yang dibutuhkan sebagai atribut untuk perhitungan penelitian akan disajikan dalam bentuk tabel dibawah ini:

Tabel 3.1 Atribut Yang Digunakan

No	Atribut	Keterangan
1	Nim	Dataset tahun 2016 sampai tahun 2018
2	Usia	Usia Mahasiswa saat mendaftar
3	Jenis Kelamin	Laki-Laki/Perempuan
4	Nilai IPK	0,00 - 4,00
5	Status Mahasiswa	Beasiswa/Regular
6	Jumlah SKS Lulus	0-24 SKS
7	Status Mahasiswa	Ikadin/Reguler
8	Status Pernikahan	Sudah menikah/Tidak menikah
9	Status Pekerjaan	Kerja/Tidak Bekerja
10	Asal Perguruan Tinggi	Unpam/Bukan Unpam
11	Klasifikasi	Lulus Tepat Waktu / Tidak Tepat Waktu

3.2 Analisis Kebutuhan Perangkat Lunak (software)

Perangkat lunak yang dibutuhkan untuk mendukung penelitian ini adalah sebagai berikut:

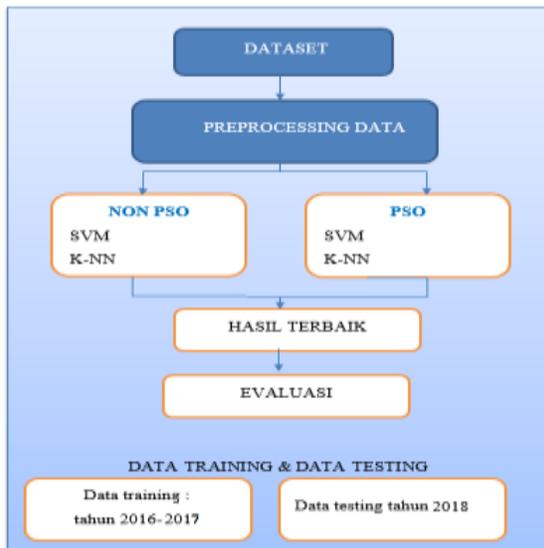
Tabel 3.2 Spesifikasi Software

No	Nama	Jenis
1	<i>Prosesor</i>	<i>Dual core ke atas</i>
2	<i>Memory</i>	4 Gb ke atas
3	<i>Harddisk</i>	Minimal 20 Gb
4	<i>Monitor</i>	14 inci
5	Lain-lain	<i>Keyboard,mouse,cdRrom,Wifi dll</i>

3.3 Analisis Kebutuhan Perangkat Keras (hardware)

Pada penelitian ini dilakukan eksperimen dengan menggunakan komputer (PC)/laptop untuk melakukan proses penghitungan terhadap model yang diusulkan. Spesifikasi perangkat keras dan sistem operasi yang digunakan pada penelitian ini adalah sebagai berikut:

Tabel 3.3 Spesifikasi Hardware



Gambar 3.1 Teknis Analisis

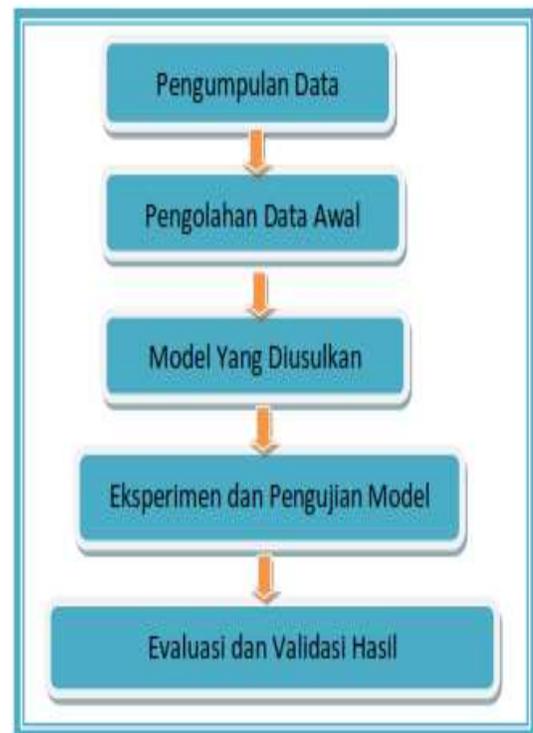
Analisis ini ditujukan untuk mengetahui algoritma dan pendekatan mana yang terbaik untuk mengurangi pengaruh ketidakseimbangan kelas, sehingga keakuratan prediksi kelas minoritas dapat ditingkatkan. Ukuran kinerja yang akan dianalisis adalah akurasi, presisi, recall dan AUC (Area Under Curve).

No	Software	Version
1	<i>Weka</i>	3.8
2	<i>Sistem Operasi</i>	<i>Microsoft Windows 7</i>

3.4 Teknis Analisis

3.5 Tahapan Penelitian

Tahapan penelitian ini bertujuan untuk mendapatkan prediksi kelulusan tepat waktu mahasiswa magister komputer STMIK Eresha, maka penelitian yang diakui/diterima harus mengikuti aturan yang berlaku. Secara umum pada penelitian ini dilakukan dengan mengikuti tahapan seperti gambar 3.2



Gambar 3.2 Tahapan Penelitian

4. HASIL DAN PEMBAHASAN

Pada subbab hasil akan dikemukakan hasil penggunaan aplikasi WEKA untuk prediksi kelulusan tepat waktu menggunakan algoritma SVM dan K-NN berbasis PSO. Pada bagian hasil prediksi dari aplikasi WEKA akan dibahas perbandingan dengan data yang sesungguhnya untuk menunjukkan bahwa aplikasi yang digunakan telah bekerja sesuai yang diharapkan. Pada subbab berikutnya akan dibahas penggunaan aplikasi untuk menghitung prediksi dengan algoritma klasifikasi SVM dan K-NN berbasis PSO. Hasil akurasi dari setiap algoritma akan dibandingkan untuk menjawab rumusan masalah.

4.1 . Pembahasan

Data penelitian yang digunakan adalah data mahasiswa magister komputer STMIK Eresha tahun 2016 sampai tahun 2018. Adapun data yang digunakan meliputi nim, usia, jenis kelamin, nilai indek prestasi sementara, sks yang di tempuh, lama studi, status mahasiswa, status pernikahan, status pekerjaan, asal perguruan tinggi.

4.2 Data Pre-Processing/Clearing

Tahap ini merupakan proses pembersihan terhadap data untuk memastikan data yang diperoleh sebelumnya dapat digunakan serta bebas dari duplikasi, kesalahan dan *validation rules* sudah sesuai. Dari *dataset* mahasiswa sebanyak 254 setelah dilakukan data *clearing* menghasilkan data sebanyak 154 data mahasiswa yang sesuai dengan kebutuhan prediksi kelulusan. Untuk langkah selanjutnya adalah *data selection*.

4.3. Pengolahan Data Awal

Data yang didapatkan kemudian diolah untuk mengurangi data yang tidak relevan, atau data dengan atribut yang hilang. Pengolahan data juga berupa konversi nilai-nilai *redundan* (berlebihan), atau nilai yang terlalu beragam ke dalam kelompok yang lebih kecil untuk mempermudah pembentukan model. Padatahap pengumpulan *dataset* yang digunakan pada Komputer tahun 2016–2018 STMIK Eresha . Data asal tersebut dilakukan *preprocessing* dengan mengisi data yang kosong (Indek Prestasi Semester 1 sampai 4) akibat tidak aktif kuliah (cuti kuliah) atau tidak kuliah selama semester itu dan hanya

membayar, sehingga datanya *blank* (kosong). Data yang kosong tersebut diisi dengan nilai minimal dari masing-masing atribut yang sama. Data yang digunakan yaitu data mahasiswa dengan variabel biodata mahasiswa, nilai mahasiswa dan keterangan lulus dan tidak lulus dari STMIK Eresha.

4.4 Klasifikasi Algoritma *Support Vector Machine*

Dalam melakukan prediksi masa depan, SVM melakukan klasifikasi data berdasarkan data *historis* dengan *scenario* yang telah diberikan. Proses pengklasifikasian ini membagi data sesuai dengan kelas lulus tepat waktu dan tidak tepat waktu, oleh karena itu data dibagi menjadi dua yaitu data *training* dan data *testing*. SVM mencari *hyperplane* (garis pemisah) terbaik dari sekian banyak kemungkinan untuk memisahkan antar kelas tersebut. *Hyperplane* yang terbaik juga dilihat dari *margin* (jarak) terbesar antara titik maksimal dari masing - masing kelas dengan *hyperplane*. Agar dapat diklasifikasi, maka data direpresentasikan sebagai *vector* berupa pasangan data (x, y). Nilai x merupakan nilai rasio kelulusan tepat waktu. Nilai *class* lulus tepat waktu diberi label 1 dan *class* tidak tepat waktu diberi label -1. Setelah dilakukan *preprocessing* data maka didapat hasil *dataset* mahasiswa magister komputer tahun 2016 yang sudah lulus semua, data ini akan digunakan sebagai data training untuk prediksi kelulusan tepat waktu tahun 2017 dan 2018. Untuk prediksi digunakan aplikasi Weka versi 3.8 dengan algoritma *Support Vector Machine*. Berikut hasil perhitungan nilai akurasi data tahun 2016 yang akan kita gunakan sebagai data *training*.

Tabel 4.2 *Confousion Matrix* Prediksi Kelulusan Algoritma SVM

Prediksi	Ya	Tidak
LulusTepatWaktu	38	3
TidakTepatWaktu	22	31

Dari tabel di atas dapat dihitung dengan rumus akurasi sebagai berikut:

$$\text{Akurasi} = \frac{(TN+TP)}{(TN+FN+TP+FP)}$$

$$\text{Akurasi} = \frac{(3+24)}{(3+28+24+4)} = \frac{(27)}{(59)} = 0,8813$$

Didapat nilai akurasi SVM sebesar $0,8813 = 88,13\%$

$$\text{Nilai precision adalah} = \frac{(TP)}{(TP+FP)} = \frac{(24)}{(24+4)} = \frac{(24)}{(28)}$$

$$= 0,8571 = 85,71 \%$$

$$\text{Nilai recall adalah} = \frac{(TP)}{(TP+TN)} = \frac{(24)}{(24+3)} = \frac{(24)}{(27)} = 0,889$$

$$= 88,9\%$$

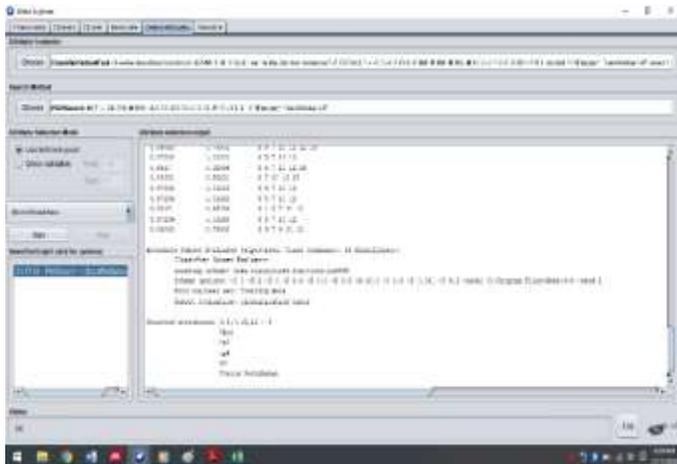
Untuk semua algoritma klasifikasi berusaha

membentuk model yang mempunyai akurasi tinggi (laju *error* rendah). Umumnya, model yang dibangun dapat diprediksi dengan benar pada semua data yang menjadi data latihnya, tetapi ketika model berhadapan dengan data uji barulah kinerja model dari sebuah algoritma klasifikasi ditentukan

4.5 Algoritma SVM Dan K-NN Berbasis PSO

4.5.1 Algoritma SVM Berbasis PSO

Setelah kita dapatkan hasil dari perhitungan *Support Vector Machine* dan *K-Nearest Neighbour* maka selanjutnya kita akan optimalkan menggunakan *Particle Swarm Optimization*.



Gambar 4.3 Algoritma SVM Berbasis PSO

Dari hasil penambahan algoritma PSO pada SVM hasil yang kita dapatkan adalah pembobotan pada setiap atribut yang digunakan untuk prediksi tepat waktu. Menggunakan *Support Vector Machine* berbasis *Particle Swarm Optimization* pada tools WEKA atribut yang berpengaruh besar terhadap prediksi lulus tepat waktu dan tidaknya kelulusan yaitu berdasarkan “*Umur, ip2, ip4, s3, Status pernikahan*”.

4.5.2. Algoritma K-NN Berbasis PSO

Setelah kita dapatkan hasil dari perhitungan *K-Nearest Neighbour* maka selanjutnya kita akan optimalkan menggunakan *Particle Swarm Optimization*. Dengan nilai k dari masing-masing prediksi mulai dari k-3, k-5, k-7, dan k-9.

a. Algoritma K-NN Dengan Nilai K=3 Berbasis PSO Setelah menggunakan *K-NN* dengan *k=3* berbasis *Particle Swarm Optimization* dan hasil yang didapatkan setelah menggunakan tools *weka*, maka ditemukanlah hal yang berpengaruh besar terhadap prediksi lulus tepat waktu dan tidaknya kelulusan yaitu berdasarkan “*Umur, ip2, ip4, sks2, sks3, status_pernikahan, status_mahasiswa*”.

b. Algoritma Algoritma K-NN Dengan Nilai K=5 Berbasis PSO

Setelah menggunakan *K-NN* dengan *k=5* berbasis *Particle Swarm Optimization* dan hasil yang didapatkan setelah menggunakan tools *weka*, maka ditemukanlah hal yang berpengaruh besar terhadap prediksi lulus tepat waktu dan tidaknya kelulusan yaitu berdasarkan “*Umur, ip4, s3, s4, status_pernikahan, Status_Mahasiswa*”.

4.6 Penentuan Algoritma Terbaik

Dari hasil prediksi dari *Support Vector Machine, K-Nearest Neighbour* dan *Support Vector Machine, K-Nearest Neighbour* berbasis *Particle Swarm Optimization* maka didapatkan algoritma terbaik untuk menyelesaikan permasalahan prediksi kelulusan tepat waktu. Hasil terbaik disajikan dalam bentuk tabel dibawah ini:

Tabel 4.5 Penentuan Algoritma Terbaik

Model	Non-				
	Svm	K=3	K=5	K=7	K=9
Akurasi	73,4 %	77,6%	76,5%	78,7%	78,73%
Presisi	91,2 %	82,0%	80,4%	81,1%	80,0%
Recall	58,5 %	77,4%	77,4%	81,1%	83,0%
Model	Penambahan algoritma PSO				
	Svm	K=3	K=5	K=7	K=9
Akurasi	73,4 1%	77,65 %	76,50%	78,73 %	78,73%
Presisi	91,2 %	82,0%	80,4%	81,1%	80,0%
Recall	58,5 %	77,4%	77,4%	81,1%	83,0%
Peningkatan	1,51 %	1,27%	1,41%	1,04%	1,48%
	74.9 1%	78.87 %	77.91%	79.74 %	80.21%

Setelah ditambahkan algoritma *Particle Swarm Optimization*, nilai akurasi tertinggi terdapat pada algoritma k-nn dengan nilai k=9 yaitu menjadi 80.21% nilai akurasinya meningkat 1.48 %.

Dari hasil penambahan algoritma *Particle Swarm Optimization* terdapat perubahan untuk nilai akurasi terbaik, yang sebelumnya akurasi terbaik didapatkan pada algoritma *K-Nearest Neighbour* dengan nilai k-7 setelah ditambahkan PSO nilai akurasi terbaik terdapat pada algoritma *K-Nearest Neighbour* dengan nilai k-9.

Dari model yang diusulkan algoritma terbaik terdapat pada K-NN dengan K-9 berbasis *Particle Swarm Optimization* dengan nilai akurasi 80,21 %,

tetapi hasil ini masih jauh dari nilai *excellent* (sangat baik).

5. KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil dan perbandingan yang telah dilakukan, kesimpulan yang bisa ditarik dari penelitian ini adalah:

1. Metode *K-Nearest Neighbour* dengan nilai $k=7$ memiliki tingkat akurasi yang lebih tinggi dibandingkan algoritma *Support Vector Machine* dan *K-Nearest Neighbour* dengan nilai k_3, k_5, k_7, k_9 . Algoritma $k=7$ terbaik dengan nilai akurasi sebesar 78,73%, dengan validasi *cross validation* 10, sedangkan nilai akurasi terendah terdapat pada algoritma *Support Vector Machine* yaitu sebesar 73,41%.
2. Setelah ditambahkan algoritma *Particle Swarm Optimization*, nilai akurasi tertinggi terdapat pada algoritma $k=9$ yaitu menjadi 80,21%, nilai akurasinya meningkat 1,48 %
3. Ketika algoritma PSO ditambahkan atribut yang memiliki bobot paling berpengaruh terhadap hasil prediksi kelulusan tepat waktu yaitu “nilai indek prestasi semester 2 dan 3, SKS semester 3 dan status Pernikahan”.
4. Hasil –hasil ini menunjukkan bahwa penambahan algoritma *K-Nearest Neighbour* memiliki performa yang lebih baik dibandingkan *Support Vector Machine* untuk masalah prediksi kelulusan tepat waktu.

5.2 Saran

Pada penelitian ini digunakan atribut IPK (Indek Prestasi Semetara), SKS (Satuan Kredit Semester) lulus, lama studi dan biodata mahasiswa hasil yang didapat masih jauh dari kata baik, maka disarankan untuk penelitian selanjutnya supaya:

1. Untuk penelitian lanjutan supaya lebih banyak lagi atribut yang digunakan, misalnya terkait mata kuliah, kemampuan ekonomi, pekerjaan orang tua dan lain-lain agar hasil yang didapat lebih maksimal.
2. Dari hasil wawancara terhadap mahasiswa yang lulus tidak tepat waktu terkait hasil prediksi yang didapat pada penelitian ini:
 - a. Tidak adanya ruang khusus bagi mahasiswa magister komputer untuk berdiskusi baik

dengan dosen maupun sesama mahasiswa

- b. Pada semester akhir atau semester 4 supaya diisi matakuliah lain selain tesis, bertujuan agar komunikasi antar mahasiswa dan dosen terpantau.
 - c. Kesibukan dalam hal sebagai status suami/istri dan karyawan yang terkadang dinas luar kota dalam waktu terbilang lama yaitu 1 bulan atau lebih menjadi faktor *external* paling berpengaruh dalam mahasiswa tidak lulus tepat waktu.
3. Untuk STMIK Eresha selaku penyelenggara IKADIN agar dapat lebih selektif memilih mahasiswa magister komputer terkait usia, status pernikahan, status pekerjaan.

DAFTAR PUSTAKA

- [1]. Ariwibowo, & Sasmito, A. (2013). Metode Data Mining Untuk Klasifikasi Kesetiaan Pelanggan Terhadap Merek Produk. *Seminar Nasional Sistem Informasi Indonesia*, 10(1), 535–541.
- [2]. Darmawan, A., Kustian, N., & Rahayu, W. (2018). Implementasi Data Mining Menggunakan Model SVM untuk Prediksi Kepuasan Pengunjung Taman Tabebuya. *STRING (Satuan Tulisan Riset Dan Inovasi Teknologi)*, 2(3), 299. <https://doi.org/10.30998/string.v2i3.2439>
- [3]. Fiqi Ruli, S. (2018). Prediksi Pergerakan Harga Saham dengan Metode Support Vector Machine (SVM) Menggunakan Trend Deterministic Data Preparation. *E-Proceeding of Engineering*, 5(3), 8356–8372.
- [4]. Kamagi, D. H., & Hansun, S. (2014). Implementasi Data Mining dengan Algoritma C4.5 untuk Memprediksi Tingkat Kelulusan Mahasiswa. *Jurnal ULTIMATICS*, 6(1), 15–20. <https://doi.org/10.31937/ti.v6i1.327>
- [5]. maulana. devi, kumalasari reza. (2019). Analisis Komparasi Algoritma Klasifikasi Data Mining Untuk Prediksi Status Kelulusan Mahasiswa Akademi Bina Sarana Informatika. *Jurnal Informatika Kaputama (JIK)*, Vol. 3 No. 2, Juli 2019, Juni(Semantik), 241–249.
- [6]. Mayadewi, P., & Rosely, E. (2015). Prediksi Nilai Proyek Akhir Mahasiswa Menggunakan. *Istem Informasi Indonesia*, November, 2–3.
- [7]. Muhammad, F., Ibnu Asror, S.T., M., & IndraLukmana Sardi. S.T., M. (2018). Analisis Perbandingan CPU dan GPU (CUDA) Pada Klasifikasi Data Mining dengan Menggunakan Metode K-Nearest Neighbour Kernel Algorithn. *E- Proceeding of Engineering*, 5(3), 7494–7506.
- [8]. Muri, L. P., Pramono, B., & Sari, J. Y. (2018). Prediksi tingkat penyakit demam berdarah di kota kendari menggunakan metode.

SemanTIK, 4(1), 103–112.

- [9]. Rakhman, A. (2017). MENGGUNAKAN METOD DECISION TREE BERBASIS PARTICLE SWARM OPTIMATION (PSO) Untuk membantu dalam menemukan berharga itu diperlukan teknik data mining . Data mining adalah teknik untuk menemukan dan mendeskripsikan pola-pola yang ada dalam data sebagai seb. *Smart Comp*, 6(1), 193–19

