ANALISIS KINERJA ALGORITMA *SUPPORT VECTOR MACHINE* (SVM) PADA DATA SELEKSI PENERIMA BEASISWA MENGGUNAKAN *PARTICLE SWARM OPTIMIZATION* (PSO) (STUDI KASUS: POLITEKNIK TEDC BANDUNG)

Ari Sudrajat[1)], Indra Budi[2)]
Sistem Informasi, Universitas AMIKOM Yogyakarta[1)], Universitas Indonesia[2)]
Email : arisoe@poltektedc.ac.id[1)], indra.ib.budi@gmail.com[2)]

Abstrak
Undang-undang No. 20 Tahun 2003 tentang Sistem Pendidikan Nasional, menyebutkan bahwa setiap peserta didik pada setiap satuan pendidikan berhak mendapatkan Beasiswa bagi yang berprestasi yang orang tuanya kurang mampu membiayai pendidikannya. Politeknik TEDC Bandung berusaha memberikan beasiswa pendidikan kepada calon mahasiswa baru dan mahasiswa lama yang telah dialokasikan oleh Kementerian Ristekdikti melalui Lembaga Layanan Pendidikan Tinggi (LLDIKTI) atau sebelumnya disebut Kopertis ataupun beasiswa pendidikan yang diselenggarakan oleh Pemerintah Pemprov dan Pemerintah Daerah. Namun jumlah calon mahasiswa penerima beasiswa yang mendaftar lebih banyak dibandingkan dengan jumlah alokasi yang disediakan untuk penerima beasiswa. Kenyataan di lapangan yang terjadi di Politeknik TEDC Bandung, calon mahasiswa yang tidak diterima sebagai penerima beasiswa mengeluh dengan sistem seleksi penerimaan beasiswa yang bersifat tertutup sehingga menimbulkan pertanyaan dan komentar baik dari orangtua / wali mahasiswa. Ketidaktepatan penentuan penerima beasiswa menjadi masalah utama yang tidak dapat terhindarkan karena setelah dilakukan analisis terhadap data calon penerima beasiswa dan penerima beasiswa di Politeknik TEDC Bandung pada tahun 2017 terdapat sebagian data penerima beasiswa yang seharusnya tidak mendapatkan beasiswa begitu juga sebaliknya. Permasalahan ini timbul dikarenakan Politeknik TEDC Bandung belum memiliki sistem pengambilan keputusan yang tepat ataupun model prediksi untuk menentukan penerima beasiswa. Sementara untuk memudahkan dan mengetahui dalam menentukan calon mahasiswa penerima beasiswa, peneliti tertarik untuk meneliti atribut-atribut yang mempengaruhinya dengan cara teknik klasifikasi yang terdapat pada data mining. Sehingga informasi hasil dari klasifikasi dapat digunakan sebagai solusi / model dalam mendukung pengambilan keputusan dalam penentuan mahasiswa penerima beasiswa di Politeknik TEDC Bandung. Adapun beberapa faktor yang menjadi persyaratan bagi calon mahasiswa penerima beasiswa di Politeknik TEDC Bandung, yaitu sebagai penerima bantuan pemerintah, pendidikan orang tua, pekerjaan orang tua, pendapatan orang tua, pengeluaran orang tua, kepemilikan rumah tinggal, jumlah tanggungan, transportasi yang digunakan, nilai rata-rata ujian, dan prestasi non akademik. *Support Vector Machine* (SVM) merupakan salah satu metode dari *data mining* yang digunakan untuk proses klasifikasi dan regresi berdasarkan pada beberapa prediksi. Algoritma *Support Vector Machine* (SVM) termasuk ke dalam 10 besar algoritma klasifikasi terbaik. Penelitian ini bertujuan untuk mengetahui model terbaik, memprediksi tingkat akurasi seleksi penerima beasiswa dan atribut yang paling berpengaruh yang dihasilkan oleh algoritma *Support Vector Machine* (SVM) berbasis *Particle Swarm Optimization* (PSO). Hasil yang diperoleh menyatakan bahwa algoritma *Support Vector Machine* berbasis *Particle Swarm Optimization* memiliki tingkat akurasi lebih besar dibandingkan dengan algoritma *Support Vector Machine* tanpa dioptimasi sebesar 78,35 % atau memiliki selisih 2,41 % dengan waktu eksekusi lebih lama 59 detik. Namun algoritma *Support Vector Machine* berbasis *Particle Swarm Optimization* tidak dapat dipastikan secara nyata hasil akurasinya lebih baik dibandingkan algoritma yang tidak dioptimasi karena melebihi nilai ambang batas yang telah ditentukan sebesar 0,168. Atribut paling berpengaruh dalam menentukan seleksi penerima beasiswa adalah atribut pengeluaran per bulan, pendidikan terakhir ayah, pendidikan terakhir ibu dan jumlah tanggungan dari total 10 atribut yang digunakan dalam proses pengujian. Dengan adanya penelitian ini diharapkan dapat membantu Politeknik TEDC Bandung dalam menentukan seleksi penerima beasiswa.

Kata Kunci: seleksi beasiswa, *Support Vector Machine* , *Particle Swarm Optimization*

Abstract
Law Number 20 of 2003 concerning the National Education System, states that every student in each educational unit has the right for a Scholarship for those who excel whose parents are less able to finance their education. Politeknik TEDC Bandung seeks to provide educational scholarships to prospective new students and old students who have been allocated by the Ministry of Research and Technology through the Higher Education Service Institute (LLDIKTI) or previously called Kopertis or educational scholarships organized by the Provincial and Local Governments. But the number of prospective scholarship recipients who register is more than the number of allocations provided for scholarship recipients. The reality on the ground that occurred at Politeknik TEDC Bandung, prospective students who were not accepted as scholarship recipients complained that the scholarship admission selection system was closed so that it

raised questions and comments from parents. Inaccuracy in determining scholarship recipients is the main problem that cannot be avoided because after analyzing data on prospective scholarship recipients and recipients of scholarships at Politeknik TEDC Bandung in 2017 there are some scholarship recipients who should not get scholarships as well as vice versa. This problem arises because Politeknik TEDC Bandung does not yet have the right decision-making system or prediction model to determine scholarship recipients. While to facilitate and find out in determining prospective students who receive scholarships, researchers are interested in examining the attributes that influence it by means of the classification techniques contained in data mining. So that the information from the classification can be used as a solution / model in supporting decision making in determining students who receive scholarships at Politeknik TEDC Bandung.The factors of students who are students receiving scholarships at Politeknik TEDC Bandung, namely as recipients of government assistance, parents education, parents job, parents income, parents outcome, homes, number of dependents, transportation, average examinations, and non academic achievement. Support Vector Machine (SVM) is one method of data mining used for processes and regression based on several predictions. Support Vector Machine (SVM) algorithm is included in the top 10 best algorithms. This reasearch aims to determine the best model, the prediction of accuracy, and others produced by the Support Vector Machine (SVM) algorithm based on Particle Swarm Optimization (PSO). The results obtained clearly that the Support Vector Machine algorithm based on Particle Swarm Optimization has a greater degree of accuracy than the Support Vector Machine algorithm without being optimized for 78.35% or has a difference 2.41% with execution time 59 seconds. But the Support Vector Machine algorithm based on Particle Swarm Optimization cannot be ascertained correctly its accuracy is better than algorithms that are not optimized because it exceeds the set limit 0.168. The deepest attributes in determining scholarship recipients are outcome in month, education parents, and the number of dependents of the total 10 attributes in the testing process. With this research, it is hoped that it can help Politeknik TEDC Banding in determining scholarship recipients

Keywords: *scholarship selection, Support Vector Machine , Particle Swarm Optimization*

I. PENDAHULUAN

Menurut Undang-undang No. 12 Tahun 2012, Pendidikan Tinggi adalah jenjang pendidikan setelah pendidikan menengah yang mencakup program diploma, program sarjana, program magister, program doktor, dan program profesi, serta program spesialis, yang diselenggarakan oleh perguruan tinggi berdasarkan kebudayaan bangsa Indonesia. Seorang lulusan SMA/SMK/MA yang akan melanjutkan ke jenjang pendidikan tinggi baik perguruan tinggi negeri maupun swasta, setidaknya membutuhkan biaya pendidikan yang tidak murah terutama bagi seorang lulusan yang memiliki keterbatasan biaya namun memiliki prestasi bidang akademik yang sangat baik. Untuk mengatasi hal tersebut, berdasarkan Undang-undang No. 20 Tahun 2003 tentang Sistem Pendidikan Nasional, menyebutkan bahwa setiap peserta didik pada setiap satuan pendidikan berhak mendapatkan Beasiswa bagi yang berprestasi yang orang tuanya kurang mampu membiayai pendidikannya.

Politeknik TEDC Bandung berusaha memberikan beasiswa pendidikan kepada calon mahasiswa baru dan mahasiswa lama yang telah dialokasikan oleh Kementerian Ristekdikti melalui Lembaga Layanan Pendidikan Tinggi (LLDIKTI) atau sebelumnya disebut Kopertis ataupun beasiswa pendidikan yang diselenggarakan oleh Pemerintah Pemprov dan Pemerintah Daerah. Namun jumlah calon mahasiswa penerima beasiswa yang mendaftar lebih banyak dibandingkan dengan jumlah alokasi yang disediakan untuk penerima beasiswa. Kenyataan di lapangan yang terjadi di Politeknik TEDC Bandung, calon mahasiswa yang tidak diterima sebagai penerima beasiswa mengeluh dengan sistem seleksi penerimaan beasiswa yang bersifat tertutup sehingga menimbulkan pertanyaan dan komentar baik dari orangtua / wali mahasiswa. Ketidaktepatan penentuan penerima beasiswa menjadi masalah utama yang tidak dapat terhindarkan karena setelah dilakukan analisis terhadap data calon penerima beasiswa dan penerima beasiswa di Politeknik TEDC Bandung pada tahun 2017 terdapat sebagian data penerima beasiswa yang seharusnya tidak mendapatkan beasiswa begitu juga sebaliknya.

Permasalahan ini timbul dikarenakan Politeknik TEDC Bandung belum memiliki sistem pengambilan keputusan yang tepat ataupun model prediksi untuk menentukan penerima beasiswa. Sementara untuk memudahkan dan mengetahui dalam menentukan calon mahasiswa penerima beasiswa, peneliti tertarik untuk meneliti atribut-atribut yang mempengaruhinya dengan cara teknik klasifikasi yang terdapat pada data mining. Sehingga informasi hasil dari klasifikasi dapat digunakan sebagai solusi / model dalam mendukung pengambilan keputusan dalam penentuan mahasiswa penerima beasiswa di Politeknik TEDC Bandung. Adapun beberapa faktor yang menjadi persyaratan bagi calon mahasiswa penerima beasiswa di Politeknik TEDC Bandung antara lain : (1) Sebagai Penerima Bantuan Pemerintah, (2) Pendidikan Orang Tua, (3) Pekerjaan Orang Tua, (4) Pendapatan Orang Tua, (5) Pengeluaran Orang Tua, (6) Kepemilikan Rumah Tinggal, (7) Jumlah Tanggungan, (8) Transportasi yang digunakan, (9) Nilai Rata-rata Ujian, dan (10) Prestasi Non Akademik. Apabila seorang pendaftar memenuhi kriteria diatas maka kemungkinan diterima menjadi mahasiswa penerima beasiswa di Politeknik TEDC Bandung.

Dalam *data mining* teknik klasifikasi merupakan proses penempatan atau pemilihan atribut dan parameter yang tepat berdasarkan dataset yang akan diklasifikasikan dalam menentukan akurasi pada proses penghitungannya. Salah satu metode teknik klasifikasi diantaranya *Support Vector Machine* dan *Decision Tree Algorithm* yang dapat digunakan untuk menentukan suatu hasil yang lebih tepat dan akurat (Somvanshi & Chavan, 2016). Klasifikasi merupakan proses penemuan suatu pola dari banyak kumpulan pola atau fungsi yang mendeskripsikan dan memisahkan kelas data satu dengan lainnya, untuk dapat digunakan untuk memprediksi data yang belum memiliki kelas data tertentu.

Support Vector Machine (SVM) merupakan salah satu metode dari *data mining* yang digunakan untuk proses klasifikasi dan regresi berdasarkan pada beberapa prediksi (Somvanshi & Chavan, 2016). Prinsip dasar *Support Vector* adalah *linear classifer* yang dapat bekerja pada masalah non-linear dengan penggunaan fungsi kernel (Hu & Gong, 2013). Pada penelitian sebelumnya yang telah dilakukan oleh Madan Somvanshi dan Pranjali Chavan dengan judul "*Review of Machine Learning Techniques Using Decision Tree and Support Vector Machine*" menyatakan bahwa penggunaan metode *Support Vector Machine* dalam proses pengklasifikasian adalah cara yang sangat cocok terutama dalam pemilihan parameter yang tepat dengan fungsi kernel untuk memberikan hasil akurasi yang lebih akurat dibandingkan dengan *Neural Network*.

Sedangkan pada peneltian lainnya yang berjudul "*Predicting Breast Cancer Recurrence Using Effective Classification and Feature Selection Technique*", peneliti menggunakan perbandigan 3 metode klasifikasi untuk mengukur tingkat akurasi yaitu SVM*, Naïve Bayes* dan *Decision Tree C4.5* dalam memprediksi kanker. Hasil menunjukan bahwa dari ketiga metode klasifikasi, SVM memiliki tingkat akurasi yang lebih tinggi dari metode yang lain sebesar 75.75%, diikuti oleh *Decision Tree* C.45 sebesar 73.73% dan *Naïve Bayes* sebesar 67.17% (Pritom, Sabab, Munshi, & Shihab, 2016). S Serta disebutkan pada bukunya Xindong Wu dan Vipin Kumar yang berjudul "The Top Ten Algorithms in Data Mining" algoritma SVM merupakan 10 besar algoritma terbaik selain Decision Tree C4.5, K-Means, Apriori, EM (expectation-maximization), Page Rank, AdaBoost, k-Nearest Neighbors, Naïve Bayes, dan CART (Classification and Regression Trees) (Wu & Kumar, 2009).

II. LANDASAN TEORI

Data mining

Data mining merupakan suatu langkah dalam menemukan pola, wawasan yang menarik serta model deskiptif untuk dimengerti dan diprediksi berdasarkan dari data yang berskala besar (Zaki & Meira, 2014). *Data mining* juga dapat dilihat sebagai hasil evolusi dari perkembangan teknologi informasi yang meliputi pengumpulan data (data collection), pembuatan basis data (database creation), pengelolaan data (data management) dan analisis data lanjutan dengan melibatkan data warehouse dan *data mining* (Han, Kamber, & Pei, 2012). Beberapa peneliti lainnya menjelaskan bahwa *data mining* sebagai proses pengambilan informasi yang tidak diketahui atau tersembunyi dari data dengan jumlah yang besar serta dapat menghasilkan potensi atau pengetahuan yang bermanfaat (Kusrini & Lutfhi, 2009).

Teknik *data mining* digunakan untuk memeriksa basis data yang berukuran besar sebagai cara untuk menemukan pola dan menjadikan sebuah informasi yang baru. Tidak semua kegiatan pencarian suatu informasi dinyatakan sebagai *Data mining* (Hermawati, 2009).

Data mining memiliki karakteristik sebagai berikut (Kusrini & Lutfhi, 2009) :

1. *Data mining* merupakan suatu proses otomatis terhadap data yang sudah ada dapat berasal dari basis data ataupun eksperimen lainnya.
2. Data yang akan diolah oleh *data mining* berupa data yang sangat besar agar hasil pengolahan dapat dijadikan sebagai rujukan yang terpercaya.
3. *Data mining* bertujuan untuk mendapatkan hubungan atau pola yang berpotensi menghasilkan pengetahuan yang bermanfaat serta sebagai pendukung pengambilan keputusan.

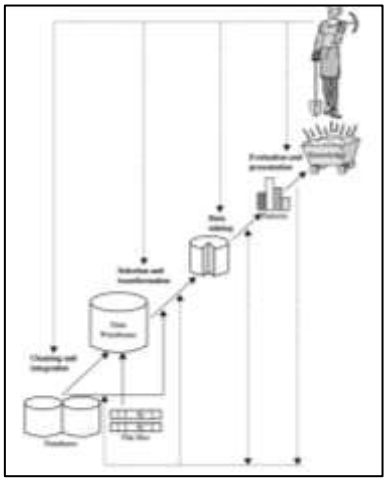
Dengan menggunakan *data mining* juga dapat memprediksikan sesuatu yang akan terjadi kemudian hari, sehingga dapat membuat pengambil kebijakan akan lebih efektif dan tepat.

Proses *Data mining*

Menurut Fajar Astuti Hermawati, tahapan atau langkah proses *data mining* merupakan bagian dari *Knowledge Discovery in Database* (KDD) (Hermawati, 2009). *Data mining* juga disebut *Knowledge Discovery in* Database (KDD) sebagai proses penemuan pengetahuan (knowledge) yang ditunjukan dalam pada gambar 1 yang bersifat iteratif (Han, Kamber, & Pei, 2012).

Tahapan atau langkah proses *data mining* ada 7 (tujuh) yaitu: pembersihan data (*data cleaning*), integrasi data (*data integration*), seleksi data (*data selection*), transformasi data (*data transformation*), *data mining*, evaluasi pola (*pattern evaluation*), dan presentasi pengetahuan (*knowledge presentation*).

Menurut Daniel T. Larose, *data mining* terdiri dari beberapa kelompok berdasarkan tugas dan fungsinya, antara lain deskripsi, estimasi, prediksi, klasifikasi, klaster dan asosiasi (Kusrini & Lutfhi, 2009).



Gambar 1. Tahapan *Data mining*

Teknik Klasifkasi

Klasifikasi merupakan suatu proses untuk menilai objek data untuk dimasukkan ke dalam kelas tertentu dari sejumlah kelas yang tersedia. Dalam klasifikasi, terdapat dua pekerjaan utama yang dilakukan, yaitu pembangunan model sebagai *prototype* untuk disimpan sebagai memori dan penggunaan model tersebut untuk melakukan pengenalan/ klasifikasi/ prediksi pada suatu objek data lain agar diketahui di kelas mana objek data tersebut dalam model yang sudah disimpannya (Hiroyasu, Obori, & Yokouchi, 2012).

Pada dasarnya algoritma *data mining* dibagi menjadi *supervised learning*, *unsupervised learning* dan *semi supervised learning*. Klasifikasi adalah contoh umum dari beberapa pendekatan *supervised learning* yang dimana satu set data yang diberikan dibagi menjadi 3 bagian yaitu *training*, validasi dan testing data set dengan label *class* yang diketahui.

Gundecha dan Liu pada jurnal penelitiannya Liza Wikarsa dan Sherly Novianti Thahir mengatakan bahwa algoritma supervised learning akan membentuk suatu model klasifikasi dari *data training* dan menggunakan model tertentu untuk memprediksinya. Untuk mengevaluasi kinerja dari model klasifikasi, model harus diterapkan pada data uji (*testing*) utnuk mendapatkan akurasi dari hasil klasifikasi (Hiroyasu, Obori, & Yokouchi, 2012).

Teknik klasifikasi dapat digunakan untuk mengelompokkan data-data yang jumlahnya besar menjadi data-data yang lebih kecil. (Calis & Boyaci, 2015).

Support Vector Machine

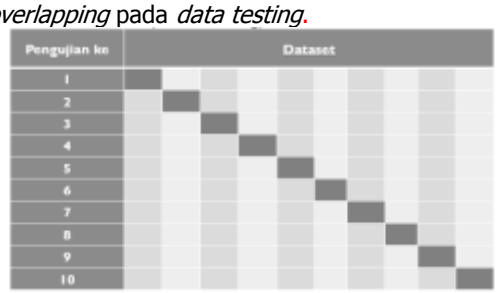
Metode klasifikasi yang kini banyak dikembangkan dan diterapkan adalah *Support Vector Machine* (SVM). SVM bekerja dengan baik pada set data dengan dimensi yang tinggi, bahkan SVM sangat baik dalam kemampuan generalisasi serta dapat bekerja pada data yang non-linear dengan teknik kernel yang dimilikinya. (Hiroyasu, Obori, & Yokouchi, 2012). SVM digunakan untuk teknik klasifikasi dan regresi yang memiliki konsep untuk mencari hyperplane terbaik yang berfungsi sebagai pemisah dua buah class pada input space.

Peneliti lain mengungkapkan bahwa SVM merupakan metode klasifikasi yang berkerja dengan cara mencari *hyperplane* dengan margin terbesar. *Hyperplane* adalah garis batas pemisah data antar kelas, sedangkan margin adalah jarak antara *hyperplane* dengan data terdekat pada masing-masing kelas. Adapun data terdekat dengan hyperplane pada masing-masing kelas inilah yang disebut *support vector* (Hu & Gong, 2013).

K-fold cross validation

Cross Validation merupakan salah satu teknik untuk menilai/memvalidasi keakuratan sebuah model yang dibangun berdasarkan dataset tertentu . Pembuatan model biasanya bertujuan untuk melakukan prediksi maupun klasifikasi terhadap suatu data baru yang boleh jadi belum pernah muncul di dalam dataset. Data yang digunakan dalam proses pembangunan model disebut data latih/training, sedangkan data yang akan digunakan untuk memvalidasi model disebut sebagai data test (Gonglong Duan and et al, 2012).

Salah satu metode *cross-validation* yang populer adalah *K-Fold Cross Validation*. Dalam teknik ini dataset dibagi menjadi sejumlah K-buah partisi secara acak. Kemudian dilakukan sejumlah K-kali eksperimen, dimana masing-masing eksperimen menggunakan data partisi ke-K sebagai *data testing* dan memanfaatkan sisa partisi lainnya sebagai *data training*. Metode *cross-validation* digunakan untuk menghindari *overlapping* pada *data testing*.



Gambar 2. Pengujian *Cross Validation*

Dari gambar 2 bahwa bisa di jelaskan warna merah yang berada pada data set merupakan *data testing* dan selebihnya merupakan *data training*. cross validation merupakan metode evaluasi standar yaitu menggunakan 10 kali pengujian. Hasil dari penilitian beserta teori menunjukan bahwa untuk mendapatkan hasil validasi yang akurat mengunakan 10-*fold cross-validation* (pengujian 10 kali menggunakan *cross validation*). 10-*fold cross-validation* akan mengulang pengujian sebanyak 10 kali dan hasil pengukuran adalah nilai rata-rata dari 10 kali pengujian (Nurhayati, dkk, 2015).

Confusion Matrix

Confusion Matrix merupakan visualisasi untuk mengevaluasi dari kinerja model klasifikasi (Herawati, 2013). Untuk melakukan klasifikasi evaluasi komparatif, maka dalam penelitian ini menggunakan *Confusion Matrix*. *Confusion Matrix* ini meliputi informasi tentang kelas yang sebenarnya dan kelas prediksi. Hal ini akan ditemukan pada kolom matriks yang mewakili kelas yang diprediksi, sedangkan setiap baris mewakili kejadian pada kelas tersebut.

Confusion Matrix adalah salah satu alat ukut berbentuk matrik 2x2 yang digunakan untuk mendapatkan jumlah ketepatan algoritma yang dipakai. *Confusion Matrix* disajikan pada tabel di bawah ini (Widodo, Handayanto , Herawati, 2013) :

Tabel 1. *Confusion Matrix*

		Actual Class	
		Class= Yes	*Class= No*
Predicted Class	*Class =Yes*	*TP*	*FP*
	Class =No	*FN*	*TN*

Dari tabel diatas menerangkan bahwa, jika yang diprediksi bernilai positif dan sesuai dengan nilai aktual (positif), maka hasilnya dikatakan *True Positive*, namun jika tidak sesuai dengan nilai aktual, maka dikatakan *False Positive* (FP). Dan sebaliknya jika yang diprediksi bernilai negatif dan aktualnya positif, maka dikatakan *False Negative* (FN), dan jika benar antara prediksi negatif dan kenyataannnya negatif, maka dikatakan *True Negative* (TN).

Informasi di dalam *confusion matrix* diperlukan untuk menentukan kinerja model klasifikasi. Ringkasan informasi ini berupa sebuah nilai yang digunakan untuk membandingkan kinerja dari model-model yang berbeda. Hal ini dapat dilakukan dengan menggunakan *performance metric*, salah satunya seperti akurasi, yang didefiniskan sebagai berikut:

$$accuracy = \frac{Number\ of\ Correct}{Total\ Number\ of\ Prediction} x\ 100\% \text{ .....(1)}$$

Particle Swarm Optimization

Particle Swarm Optimization (PSO) merupakan teknik optimasi yang berbasis populasi yang di usulkankan oleh Eberhart dan Kennedy pada tahun 1995. Metode ini terinspirasi pada prilaku sosial sekawanan burung dan ikan (*swarm*). Dalam mencari solusi yang optimal, partikel tersebut bergerak pada arah yang terbaik dari sebelumnya, yakni posisi terbaik secara global. Seperti Ke-i dinyatakan $X_i=[X_{i1}, X_{i2,....} X_{iD}]$ dalam ruang dimensi. (Bisilisin, dkk, 2014).

Pada algoritma PSO, vector kecepatan diperbaharui untuk masing-masing partikel, kemudian menjumlahkan vector kecepatan tersebut ke posisi partikel. Proses memperbaharui kecepatan dipengaruhi oleh kedua solusi yaitu melakukan penyesuaian posisi terbaik dari partikel (*particle best)* dan penyesuaian terhadap partikel terbaik dari seluruh kumpulan (*global best*). Pada tiap iterasi, setiap solusi yang direpresentasikan oleh posisi partikel dievaluasi dengan cara memasukkan solusi tersebut ke dalam *fitness function.* Prosedur algoritma PSO (Bisilisin, dkk, 2014) adalah :

1. Inisialisasi populasi dari partikel-partikel dengan posisi dan kecepatan secara acak,
2. Evaluasi nilai *fitness* untuk masing-masing partikel.
3. Pembandingan dan pembaharuan *particle best* dan *global best* untuk setiap partikel berdasarkan *fitness function.*

Proses perhitungan kecepatan dan pembaharuan partikel dihitung menggunakan persamaan di bawah ini.

$$v_i(t) = wv_i(t-1) + c_1 r_1 (x_{pi} - x_i) + c_2 r_2 (x_{gi} - x_i)$$
$$x_i(t) = x_i(t-1) + v_i(t)$$

dengan
i = indeks partikel
t = iterasi
w = *inertia*
v_i = kecepatan partikel ke-i
x_i= posisi partikel ke-i
x_{gi} = posisi terbaik dari semua partikel (gbest)
x_{pi} = posisi terbaik dari partikel ke-i (pbest)
$c_{1,2}$ = *learning rate*
$r_{1,2}$ = bilangan acak [0,1]

T-Test

T-Test adalah metode pengujian hipotesis dengan menggunakan satu individu (objek penelitian) dengan menggunakan dua perlakuan yang berbeda. Walaupun dengan menggunakan objek yang sama tetapi sampel tetap terbagi menjadi dua yaitu data dengan perlakuan pertama dan data dengan perlakuan kedua. *Performance* dapat diketahui dengan cara membandingkan kondisi objek penelitian pertama dan kondisi objek pada penelitian kedua (Hastuti, 2012).

Tabel 2. Keterangan Aturan Hasil Uji *T-Test*

No	Aturan	Keterangan
1	Jika Probabilitas = 0,050	Memiliki perbedaan yang signifikan
2	Jika Probabilitas > 0,050	Tidak memiliki perbedaan yang signifikan

Rapidminer

Rapidminer adalah salah satu aplikasi bersifat *opensource* yang dapat digunakan untuk melakukan proses *data mining*. Salah satu metode *data mining* adalah menggunakan regresi linier. Regresi linier merupakan metode statistik yang digunakan untuk untuk melakukan estimasi atau perkiraan berdasarkan data yang ada. *Rapidminer* adalah solusi untuk melakukan analisis terhadap *data mining*, text mining dan analisis prediksi yang menggunakan berbagai teknik deskriptif dan prediksi dalam memberikan wawasan kepada pengguna sehingga dapat membuat keputusan yang paling baik. (Muis & Affrandes, 2013).

III. METODE PENELITIAN

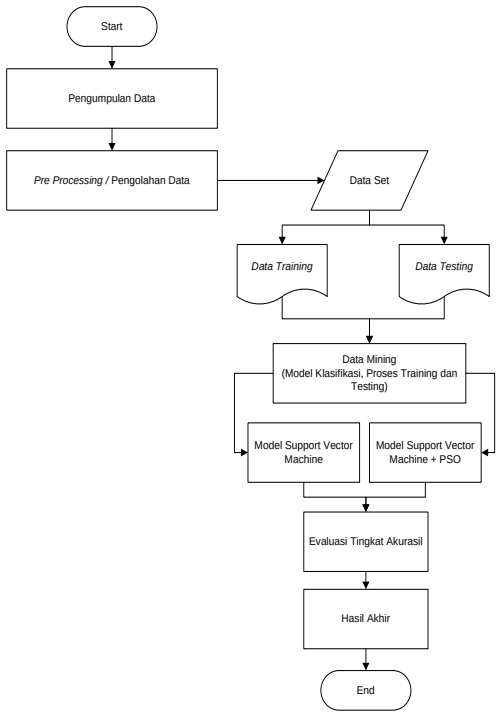
Pendekatan yang digunakan dalam penelitian ini adalah pendekatan kausal komparatif yang bertujuan untuk mengetahui atau menyelidiki kemungkinan variabel yang berpengaruh terhadap suatu objek penelitian berdasarkan atas pengamatan langsung melalui data tertentu.

Peneliti melakukan penelitian berdasarkan atribut-atribut yang memungkinkan menjadi faktor pada proses seleksi penerimaan beasiswa. Kemudian peneliti akan menghitung tingkat akurasi menggunakan algoritma *Support Vector Machine* dari masing-masing tipe kernel sebagai metode dalam memprediksi kelayakan seseorang sebagai penerima beasiswa.

Data kuesioner yang sudah terkumpul, selanjutnya akan dilakukan proses penyeleksian data, pemilihan data dan transformasi data yang akan digunakan untuk proses *data mining*. Pengolahan data awal pada kuesioner dilakukan untuk memilah atribut yang berpengaruh pada proses seleksi penerimaan beasiswa agar dapat mempercepat proses pengklasifikasian pada *Rapidminer*. Maka data yang akan diolah pada *Rapidminer* sebanyak 11 atribut dari 17 atribut yang tersedia pada kuesioner. Data yang digunakan akan dijadikan data set pada proses pengujian menggunakan algoritma *Support Vector Machine* .

Proses awal *data mining* yang dilakukan yaitu membagi data menjadi 2 bagian yang tediri dari data latih (*data training*) dan data uji (*data testing*). Pengujian data menggunakan tipe kernel algoritma *Support Vector Machine* dan *cross validation* pada perangkat lunak *Rapidminer*. Pengujian tersebut bertujuan unutk memperoleh tingkat akurasi dalam penentuan penerima beasiswa. Hasil pengujian data klasifikasi atau tingkat akurasi akan diuji kembali menggunakan *Particle Swarm Optimization* dan *T-Test* sebagai evaluasi pola yang dihasilkan oleh proses *data mining*.

Pada gambar 3, dijelaskan mengenai alur penelitian yang digunakan untuk menentukan mahasiswa penerima beasiswa menggunakan metode *Support Vector Machine* (SVM) yang digambarkan dalam bentuk diagram alir.



Gambar 3. Alur penelitian

Penjelasan alur penelitian pada **Gambar 3** adalah sebagai berikut:

1. Pengumpulan data, pengumpulan data yang dilakukan melalui studi literatur / tinjauan pustaka mengenai referensi yang berhubungan dengan metode yang akan digunakan.
2. *Pre-processing* data, penyeleksian data hasil kuesioner yang akan digunakan pada proses klasifikasi *data mining*.
3. Data set, yang terdiri atas *data training* dan *data testing* yang akan diproses dengan menggunakan algoritma *Support Vector Machine* pada tools Rapid Miner.
4. *Data mining*, pemodelan data klasifikasi melalui proses latih dan uji menggunakan algoritma *Support Vector Machine* pada tools Rapid Miner.
5. Pengujian model dilakukan dengan membandingkan algoritma *Support Vector Machine* dengan menggunakan *Particle Swarm Optimization* (PSO).
6. Evaluasi Model, mengevaluasi pola yang terbentuk dari hasil klasifikasi pada confusion matrix.

IV. HASIL DAN PEMBAHASAN

Pengolahan Data Awal

Data yang digunakan adalah data primer, yaitu data yang didapatkan langsung dari objek penelitian melalui hasil pengisian kuesioner.

Data-data yang dikumpulkan adalah data seleksi mahasiswa penerima beasiswa di Politeknik TEDC Bandung sebanyak 374 data. Data tersebut akan digunakan sebagai sampel untuk mengklasifikasi data seleksi penerima beasiswa menggunakan *Rapidminer*. Data yang dikumpulkan tidak hanya berasal dari data kuesioner melainkan ditambah dengan data manual yang tersedia di bagian kemahasiswaan Politeknik TEDC Bandung baik data seleksi beasiswa bidik misi, data seleksi beasiswa prestasi dan data seleksi beasiswa pemerintah daerah.

Data tersebut diharapkan dapat menghasilkan informasi terbaik dalam menentukan pola klasifikasi seleksi penerima beasiswa dan dapat digunakan sebagai model dalam penyeleksian penerima beasiswa di Politeknik TEDC Bandung.

Sebelum dilakukan pengolahan data menggunakan *Rapidminer*, data kuesioner yang sudah dikumpulkan harus melalui pengolahan data awal (preparation data) untuk mendapatkan data yang berkualitas dengan cara mengurangi data yang tidak relevan atau data dengan atribut yang hilang dan menghapus data-data yang bernilai sama atau redudan. Pengolahan data ini bertujuan agar pada proses pengujian menggunakan algoritma *Support Vector Machine* pada *Rapidminer* dapat dilakukan dengan mudah dan cepat serta meningkatkan akurasi. Proses preparation data dilakukan secara manual menggunakan *Microsoft Excel* dengan cara mengkoreksi ulang data kuesioner. Adapun data yang digunakan dalam penelitian ini bersifat kategorikal dan memiliki variabel atau atribut data seleksi penerima beasiswa yang dianggap penting dan berpengaruh yang disajikan pada tabel 3 untuk dijadikan sebagai data set (sampel) yang akan diuji menggunakan algoritma *Support Vector Machine* .

Tabel 3. Data Set

No	Atribut	Kategori Penilaian
1	Penerima Bantuan	1. Kartu Indonesia Pintar (KIP) 2. Bantuan Siswa Miskin (BSM) 3. Kartu Perlindungan Sosial 4. Bukan Penerima
2	Pendidikan Ayah	1. SD 2. SMP/MTs 3. SMA/SMK/MA 4. D3 5. D4/S1 6. S2
3	Pendidikan Ibu	1. SD 2. SMP/MTs 3. SMA/SMK/MA 4. D3 5. D4/S1 6. S2
4	Pendapatan Per Bulan	1. Rp. 0 - Rp. 2.000.000,- 2. Rp. 2.000.001 - Rp. 3.000.000,- 3. Rp. 3.000.001 - Rp. 4.000.000,- 4. Rp. 4.000.001 - Rp. 5.000.000,- 5. Lebih dari Rp. 5.000.000,-
5	Pengeluaran Per Bulan	1. Rp. 0 - Rp. 2.000.000,- 2. Rp. 2.000.001 - Rp. 3.000.000,- 3. Rp. 3.000.001 - Rp. 4.000.000,- 4. Rp. 4.000.001 - Rp. 5.000.000,- 5. Lebih dari Rp. 5.000.000,-
6	Kepemilikan Rumah Tinggal	1. Sendiri 2. Sewa 3. Lainnya
7	Jumlah Tanggungan	1. 1 2. 2 3. 3 4. 4 5. Lebih dari 4
8	Transportasi yang digunakan	1. Jalan Kaki 2. Sepeda 3. Transportasi Umum 4. Motor Pribadi 5. Mobil Pribadi
9	Nilai Rata-rata UNAS	(input dari responden)
10	Prestasi Non Akademik	1. Tingkat Kota/Kab 2. Tingkat Propinsi 3. Tingkat Nasional 4. Tidak Ada
11	Hasil Seleksi	1. Diterima 2. Tidak

Beberapa atribut dan kategori penilaian diatas perlu ditransformasikan agar dapat memudahkan dalam pembentukan model seleksi penerima beasiswa pada *Rapidminer*. Kategori penilaian yang ditransformasikan adalah pada atribut Bantuan Pemerintah dan atribut Prestasi Non Akademik yang disajikan pada tabel 4.

Tabel 4. Transformasi Data

No	Atribut	Kategori Penilaian	Hasil Transformasi
1	Penerima Bantuan	1. Kartu Indonesia Pintar (KIP) 2. Bantuan Siswa Miskin (BSM) 3. Kartu Perlindungan Sosial 4. Bukan Penerima	1. Penerima 2. Bukan Penerima
2	Prestasi Non Akademik	1. Tingkat Kota/Kab 2. Tingkat Propinsi 3. Tingkat Nasional 4. Tidak Ada	1. Ada 2. Tidak Ada

Pada tabel 4 menjelaskan hasil transformasi data yaitu :

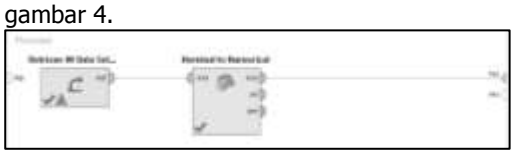
1. Atribut Penerima Bantuan, sebelumnya memiliki 4 kategori penilaian dan ditransformasikan menjadi 2 kategori penilaian yaitu Penerima dan Bukan Penerima bantuan pemerintah.
2. Atribut Prestasi Non Akademik, sebelumnya memiliki 4 kategori penilaian dan ditransformasikan menajdi 2 kategori penilaian yaitu Ada dan Tidak Ada.

Setelah melalui tahap *preparation* data dan transformasi data, maka data set seleksi beasiswa ini dapat melanjutkan ke tahap pengujian menggunakan algoritma *Support Vector Machine* pada *Rapidminer* untuk mengetahui model dengan tingkat akurasi yang dihasilkan dari algoritma tersebut.

Pemodelan *Data mining*

Dalam penelitian ini, penulis akan membangun model pengklasifikasian *data mining* yang dibagi menjadi 2 model yaitu klasifikasi menggunakan algoritma *Support Vector Machine* , optimasi klasifikasi algoritma *Support Vector Machine* menggunakan *Particle Swarm Optimization* (PSO) untuk mengetahui atribut yang paling berpengaruh pada data seleksi penerima beasiswa.

Data set yang telah diperoleh dari hasil pengisian kuesioner bersifat kategorikal, sedangkan algoritma *Support Vector Machine* hanya dapat menguji data set yang bersifat numerik. Maka sebelum dilakukan pengujian menggunakan algoritma *Support Vector Machine* , data set akan dikonversikan ke dalam bentuk numerik dengan menggunakan *blending type* yaitu *nominal to numerical* seperti tertera pada gambar 4.



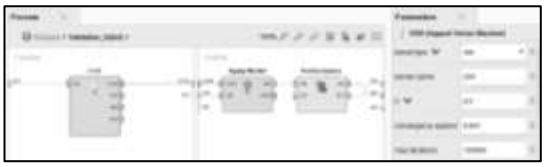
Gambar 4. Konversi data set

Pembangunan Model Pengujian SVM

Berikut pembangunan model dan pengujian algritma SVM pada tools *Rapidminer* disajikan pada gambar 5 di bawah ini.

Gambar 5. Model SVM

Di dalam blok validasi dibagi menjadi 2 bagian model yaitu model training dan model testing. Pada model training diisi oleh algoritma SVM dengan spesifikasi tipe kernel dot, parameter C = 0.00, sedangkan pada model testing diisi dengan Apply Model dan Performance untuk mengukur tingkat akurasi yang dihasilkan oleh algoritma SVM pada diagram confusion matrix.

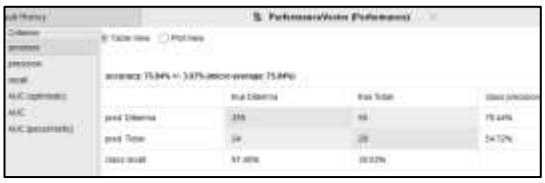


Gambar 6. Model *training* dan *testing*

Setelah dilakukan pengujian, didapatkan hasil *accuracy, precision, recall* dan *AUC* dari *performance vector* yang disajikan dalam bentuk tabel dan gambar dari *capture* program pada tabel 5 dan gambar 6.

Tabel 5. Hasil Performance Vector Algoritma SVM

Performance Vector	**Hasil**
Accuracy	75,94 %
Precision	63,01 %
Recall	30,44 %
AUC	0,787



Gambar 7. *Confusion matrix* algoritma SVM

Pada gambar 7 dijelaskan mengenai visualisasi *confusion matrix, class precision* dan *class recall*. Adapun lama waktu proses pengujian menggunakan algoritma *Support Vector Machine* memiliki waktu proses relatif singkat berkisar 1 (satu) detik. Hasil dari tabel 5, jika dilakukan perhitungan manual maka diperoleh hasil sebagai berikut :

$$accuracy = \frac{255 + 29}{374} x\ 100\% = 75.93\ \%$$

Hasil perhitungan accuracy manual di atas tidak jauh berbeda dengan hasil perhitungan pada *Rapidminer* yang bernilai 75,94 % dan hasil perhitungan manual sebesar 75,93 % atau terdapat selisih 0,01 %.

Pembangunan Model Pengujian SVM-PSO

Setelah dilakukan pengujian pad algoritma SVM selanjutnya akan dilakukan pembangunan model dan pengujian algoritma SVM menggunakan PSO pada *Rapidminer*.

Hasil pengujian sebelumnya, tingkat akurasi kinerja algoritma SVM menunjukan persentase sebesar 75,94 %. Dan pada gambar 8 ditunjukan pembangunan model dan pengujian kinerja algoritma SVM dengan menggunakan PSO.

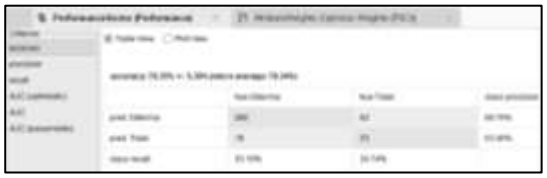
Gambar 8. Model SVM Berbasis PSO

Pada gambar 8, PSO berisi validasi yang didalamnya di bagi menjadi proses *training* terdapat algoritma SVM dan proses *testing* diisi dengan *Apply Model* dan *Performance*. Yang berbeda dengan pengujian sebelumnya, pada operator validasi yang dihubungkan ke hasil (*result*) hanya pada *performance vector* nya saja.

Berikut ini hasil dari pengujian berbasis PSO terhadap algoritma SVM . Hasil yang didapatkan terdiri dari *accuracy, precision, recall* dan *AUC* yang disajikan pada tabel 6 dan gambar 9 dengan waktu eksekusi selama 60 detik.

Tabel 6. Hasil *Performance Vector* Algoritma SVM Berbasis PSO

Performance Vector	**Hasil**
Accuracy	78,35 %
Precision	65,44 %
Recall	34,78
AUC	0,793



Gambar 9. *Confusion Matrix* Algoritma SVM Berbasis PSO

Selain nilai *accuracy, precision, recall* dan *AUC* meningkat dari hasil pengujian sebelumnya, didapatkan atribut-atribut yang paling berpengaruh hasil pengujian dengan menggunakan PSO yang disajikan pada tabel 7 beserta bobot penilaiannya.

Tabel 7. Atribut Paling Berpengaruh

Atribut	**Bobot**
Penerima Bantuan = Bukan	0,992
Penerima Bantuan = Penerima	0,487
Pendidikan Terakhir Ayah = SMA/SMK/MA	0,592
Pendidikan Terakhir Ayah = D4/S1	0,298
Pendidikan Terakhir Ayah = SD	0,635
Pendidikan Terakhir Ayah = SMP/MTs	0,664
Pendidikan Terakhir Ayah = D3	0,147
Pendidikan Terakhir Ayah = S2	0,352
Pendidikan Terakhir Ibu = SMA/SMK/MA	0,256
Pendidikan Terakhir Ibu = D4/S1	0,803
Pendidikan Terakhir Ibu = SD	0,463
Pendidikan Terakhir Ibu = SMP/MTs	0,091
Pendidikan Terakhir Ibu = S2	0,062
Pendidikan Terakhir Ibu = D3	0,693
Pendapatan Per Bulan = Rp, 2,000,001 - Rp, 3,000,000,-	0,922
Pendapatan Per Bulan = Rp, 3,000,001 - Rp, 4,000,000,-	0,018
Pendapatan Per Bulan = Lebih dari Rp, 5,000,000,-	0,223
Pendapatan Per Bulan = Rp, 0 - Rp, 2,000,000,-	0,170
Pendapatan Per Bulan = Rp, 4,000,001 - Rp, 5,000,000,-	0,420
Pengeluaran Per Bulan = Rp, 2,000,001 - Rp, 3,000,000,-	0,272
Pengeluaran Per Bulan = Rp, 3,000,001 - Rp, 4,000,000,-	0,715
Pengeluaran Per Bulan = Lebih dari Rp, 5,000,000,-	0,872
Pengeluaran Per Bulan = Rp, 4,000,001 - Rp, 5,000,000,-	0,882
Pengeluaran Per Bulan = Rp, 0 - Rp, 2,000,000,-	0,523
Kepemilikan Rumah Tinggal = Sendiri	0,574
Kepemilikan Rumah Tinggal = Lainnya	0,950
Kepemilikan Rumah Tinggal = Sewa	0,079
Jumlah Tanggungan = Lebih dari 4	0,170
Jumlah Tanggungan = 3	0,306
Jumlah Tanggungan = 4	0,389
Jumlah Tanggungan = 2	0,962
Jumlah Tanggungan = 1	0,222
Transportasi = Transportasi Umum	0,048
Transportasi = Motor Pribadi	0,981
Transportasi = Mobil Pribadi	0,036
Transportasi = Sepeda	0,206
Transportasi = Jalan Kaki	0,093
Prestasi Non Akademik = Tidak	0,871
Prestasi Non Akademik = Ada	0,268
Nilai Rerata UNAS	0,398

Atribut-atribut pada tabel 7 akan diklasifikasikan dalam 10 kelompok besar berdasarkan kelompok atributnya dan dihitung setiap bobotnya. Atribut yang dinyatakan paling berpengaruh akan memiliki nilai bobot yang lebih besar dari rata-rata seluruh atribut. Pada tabel 8 merupakan hasil pengklasifikasian dan pembobotan atribut yang paling berpengaruh.

Tabel 8. Pengklasifikasian Atribut Berbasis PSO

Klasifikasi Atribut	Total Bobot
Pengeluaran Per Bulan	3,264
Pendidikan Terakhir Ayah	2,687
Pendidikan Terakhir Ibu	2,367
Jumlah Tanggungan	2,049
Pendapatan Per Bulan	1,752
Kepemilikan Rumah Tinggal	1,604
Penerima Bantuan	1,479
Transportasi	1,363
Prestasi Non Akademik	1,139
Nilai Rerata UNAS	0,398
Total	**18,102**
Nilai Bobot rata-rata	**1,810**

Berdasarkan pembobotan pada tabel 8, maka atribut yang paling berpengaruh adalah atribut yang memiliki bobot lebih dari 1,810 yaitu atribut pengeluaran per bulan (3,264), pendidikan terakhir ayah (2,687), pendidikan terakhir ibu (2,367) dan jumlah tanggungan (2,049).

Analisis Kinerja Algoritma SVM

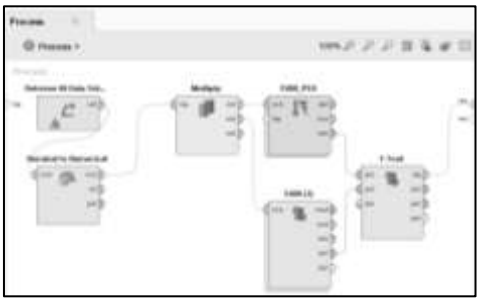
Keseluruhan pengujian telah dilakukan terhadap algoritma SVM pada data seleksi penerima beasiswa di Politeknik TEDC Bandung, baik yang sebelum dioptimasi dan setelah dioptimasi menggunakan PSO. Hasil perbandingan disajikan pada tabel 9 sebagai hasil akhir pengujian serta sebagai dasar atau pendukung dalam membuat kesimpulan hasil penelitian.

Tabel 9. Komparasi Kinerja Algortima SVM

Parameter Kinerja	Model SVM	Model SVM-PSO
Accuracy	75,94 %	78,35 %
Precision	63,01 %	65,44 %
Recall	30,44 %	34,78 %
AUC	0,787	0,793
Waktu (detik)	1	60

Dari tabel 9 di atas dapat dijelaskan bahwa dari seluruh parameter kinerja model SVM berbasis PSO mengalami peningkatan tetapi dari waktu eksekusi lebih lama 59 detik.

Guna menambah kepastian bahaw algoritma SVM berbasis PSO memiliki kinerja lebih tinggi dibandingkan dengan algortima SVM tanpa di optimasi, maka perlu dilakukan pengujian signifikansi menggunakan *T-Test* pada *Rapidminer* ditunjukan pada gambar 10.



Gambar 10. Pengujian *T-Test*

Hasil Pengujian *T-Test* berupa matrik yang ditunjukan pada gambar 11 di bawah ini.

Gambar 11. Hasil Uji *T-Test*

Standar nilai ambang batas untuk menentukan signifikansi adalah 0.05. Sehingga hasil pengujian pada gambar 11, menunjukkan bahwa nilai tingkat signifikansi pada algoritma SVM berbasis PSO menghasilkan nilai sebesar 0,168 lebih besar dari nilai ambang batas. Hal tersebut berarti bahwa tidak dapat dipastikan kinerja algoritma SVM berbasis PSO secara nyata akan lebih baik dari algoritma SVM.

V. KESIMPULAN DAN SARAN

Kesimpulan

Berdasarkan hasil penelitian dan pembahasan, maka dapat ditarik kesimpulan :

1. Pengujian pada model algoritma SVM berbasis PSO pada data seleksi penerima beasiswa menghasilkan nilai akurasi yang lebih besar dibandingkan model algoritma SVM yaitu sebesar 78,35 % atau memiliki selisih 2,41 %. Namun butuh waktu eksekusi lebih lama 59 detik.
2. Hasil pengujian signifikansi menggunakan *T-Test* didapatkan nilai sebesar 0,168 sehingga tidak dapat dipastikan secara nyata bahwa algoritma SVM berbasis PSO akan lebih baik hasil akurasinya dibandingkan algoritma yang tidak dioptimasi karena melebihi nilai ambang batas yang telah ditentukan.
3. Hasil pengujian pada data seleksi penerima beasiswa didapatkan atribut-atribut yang paling berpengaruh dalam menentukan penerima beasiswa yaitu atribut pengeluaran per bulan (3,264), pendidikan terakhir ayah (2,687), pendidikan terakhir ibu (2,367) dan

jumlah tanggungan (2,049) dari total 10 atribut yang digunakan dalam proses pengujian.

Saran

Berdasarkan hasil penelitian yang telah dilakukan penulis, penulis akan memberikan saran sebagai pertimbangan untuk melakukan pengembangan model ke depannya, yaitu :

1. Diharapkan penelitian berikutnya mampu melakukan pengembangan model dengan membandingkan hasil akurasi pada algoritma SVM dengan menentukan nilai C parameter dan tipe kernel yang terdapat pada algoritma SVM.
2. Diharapkan penelitian berikutnya mampu menggunakan data set yang lebih banyak dengan model yang berbeda sehingga dapat meningkatkan tingkat akurasi prediksi yang lebih baik.

DAFTAR PUSTAKA

Bisilisin, F.Y., Herdiyeni, Y., & Silalahi, B.P. (2014). *Optimasi K-Means Clustering Menggunakan Particle Swarm Optimization Pada Sistem Identifikasi Tumbuhan Obat Berbasis Citra*. Jurnal Ilmu Komputer Agri-Informatika, 38-47.

Calis, A., & Boyaci, A. (2015). *Data Mining Application in Banking Sector with Clustering and Classification Methods. International Conference on Industrial Engineering and Operations Management*. Dubai: IEEE Conference Publications.

Duan, G., Liu, P., Liu, R., & Wei, L. (2012). A *Selection Methods of Feature Attributes Based on RS-SVM . International Conference on Natural Computation (ICNC)* (pp. 114 - 117). IEEE Conference Publications.

Han, J., Kamber, M., Jian Pei. (2011*). Data mining: Concepts and Techniques.* 3rd Edition. San Francisco: Morgan Kaufmann Publisher.

Hastuti, K. (2012). *Analisis Komparasi Algoritma Klasifikasi Data Mining Untuk Prediksi Mahasiswa Non Aktif*. Seminar Nasional Teknologi Informasi & Komunikasi Terapan, (pp. 241 - 249). Semarang.

Hermawati, F.A. (2009). *Data Mining*. Yogyakarta : Penerbit Andi.

Hiroyasu, T., Obori, Y., & Yokouchi, H. (2012). *Classification Method into Determinable and Indeterminable Areas Using SVM and Learning Data Selection*. SCIS-ISIS (pp. 595 - 599). Kobe: IEEE Conference Publications.

Hu, L.-F., & Gong, W. (2013). *A Method For Feature Selection Based On The Optimal Hyperplane of SVM and Independent Analysis*. International Conference on Machine Learning and Cybernetics (pp. 114 - 117). Tianjin: IEEE Conference Publications.

Kusrini, & Luthfi, E. T. (2009). *Algoritma - Data Mining.* Yogyakarta : Penerbit Andi.

Kusumodestoni, H., & Sarwido. (2017). *Komparasi Model Support Vector Machine dan Neural Network Untuk Mengetahui Tingkat Akurasi Prediksi Tertinggi Saha.* Jurnal Informatika UPGRIS.

Larose, D. T. (2006). *Data mining Methods And Models.* New Jersey: John Wiley and Son, Inc.

Muis, I. A., & Affrandes, M. (2013). *Belajar Data Mining dengan Rapid Miner*. Jurnal Sains, Teknologi dan Industri, (pp. 189-197). Jakarta.

Nurhayati, S., Kusrini, & Luthfi, E. T. (2015). *Prediksi Mahasiswa Drop Out Menggunakan Support Vector Machine*. Jurnal Ilmiah Sisfotenika, 85 - 93.

Pritom, A. I., Sabab, S. A., Munshi, M. A., & Shihab, S. (2016). *Predicting Breast Cancer Recurrence Using Effective Classification and Feature Selection Technique.* International Conference on Computer and Information Technology (pp. 310 - 314). Dhaka: IEEE Conference Publications.

Saifudin, A. (2018). *Metode Data Mining Untuk Seleksi Calon Mahasiswa Pada Penerimaan Mahasiswa Baru Di Universitas Pamulang*. Jurnal Teknologi, 25-36.

Somvanshi, M., & Chavan, P. (2016). *A Review of Machine Learning Techniques Using Decision Tree and Support Vector Machine*. International Conference on Computing Communication Control and automation (ICCUBEA) (pp. 1-7). IEEE Conference Publications.

Widodo, Handayanto & Herlawati. (2013). *Penerapan Teknik Data Mining dengan Matlab*. Bandung: Rekayasa Sains.

Wu, X., & Kumar, V. (2009). *The Top Ten Algorithms in Data Mining*. Minnesota: Taylor & Francis Group.

Zaki, M. J., & Meira, W. (2014). *Data Mining and Analysis - Fundamental Concepts and Algorithms.*