

Analisis Perbandingan Kinerja Algoritma Klasifikasi Data Mining Penelusuran Minat Calon Mahasiswa

Comparative Analysis of Data Mining Classification Algorithm Performance for Searching Prospective Student Interests

¹Budiman*, ²Zatin Niqotaini

Sistem Informasi, Fakultas Teknologi dan Informatika, Universitas Informatika dan Bisnis Indonesia
Jl. Soekarno-Hatta No.643, Sukapura, Kec. Kiaracondong, Kota Bandung, Jawa Barat, Indonesia

*e-mail: budiman1982@gmail.com

(*received*: 25 November 2021 , *revised*: 25 November 2021, *accepted*: 20 Desember 2021)

Abstrak

Penerimaan mahasiswa baru merupakan kegiatan yang selalu dilaksanakan setiap Perguruan Tinggi dalam menyambut Tahun Akademik baru. Penurunan jumlah pendaftar pada setiap tahun terutama masa pandemi Covid-19 saat ini menjadi permasalahan AMIK HASS dalam perekrutan mahasiswa, maka diperlukan upaya untuk mengolah data penerimaan mahasiswa baru yang ada. Penerapan data mining menggunakan algoritma klasifikasi salah satunya bertujuan untuk melakukan prediksi. Algoritma klasifikasi yang digunakan yaitu Naïve Bayes (NB), Decision Tree J48 (J48), K-Nearest Neighbor (K-NN), Random Forest (RF) dan Support Vector Machine (SVM). Pengujian kelima algoritma klasifikasi bertujuan untuk menganalisis kinerja masing-masing algoritma, sehingga memberikan informasi dan pengetahuan data set sebagai rekomendasi dalam strategi promosi dimasa yang akan datang. Penelitian ini menggunakan Machine Learning Waikato Environment for Knowledge Analysis (WEKA). Data set yang digunakan dalam penelitian ini adalah penelusuran potensi dan minat calon mahasiswa baru sebanyak 5934 record. Pengujian yang dilakukan pada kelima algoritma klasifikasi menggunakan mode test percentage split yaitu 70% untuk data training dan 30% untuk data testing. Hasil pengujian menunjukkan tingkat akurasi tertinggi pada algoritma klasifikasi J48 sebesar 90,34% diikuti RF sebesar 89,04%, SVM sebesar 88,43%, K-NN sebesar 87,53%, SVM sebesar 87,25%. Berdasarkan hasil pengujian algoritma klasifikasi J48 merupakan algoritma terbaik terhadap pengujian data set dengan tingkat kesalahan prediksi paling rendah sebesar 0,26. Algoritma klasifikasi Decision Tree J48 memiliki aturan eksplisit untuk klasifikasi dan penanganan data heterogen yang tepat dengan memiliki rule sebanyak 255.

Kata kunci: Data Mining, Naïve Bayes, Decision Tree J48, K-Nearest Neighbor, Random Forest, Support Vector Machine

Abstract

Admission of new students is an activity that's always carried out by every university in the new academic year. The decline in the number of registrants every year is an obstacle for AMIK HASS in new student admissions, efforts are needed to process the existing data on new student admissions. Data mining applications use classification algorithms that aim to make predictions. The algorithms used are Nave Bayes (NB), Decision Tree J48 (J48), K-Nearest Neighbor (K-NN), Random Forest (RF), and Support Vector Machine (SVM). Algorithm testing to analyze the performance of each algorithm uses WEKA. The data set used in this study is the search for potential and interest of prospective new students as many as 5,934 records. The tests carried out on the five algorithms use the test percentage split mode, which is 70% for training data and 30% for test data. The highest accuracy rate on J48 is 90.34% followed by RF at 89.04%, SVM at 88.43%, K-NN at 87.53%, NB at 87.25%. J48 is the best algorithm for testing data sets with the lowest prediction error rate of 0.26. The J48 classification algorithm has explicit rules for the classification and handling of heterogeneous data by having 255 rules.

Keywords: Data Mining, Naïve Bayes, J48 Decision Tree, K-Nearest Neighbor, Random Forest, Support Vector Machine

1 Pendahuluan

Mahasiswa merupakan aset dalam sebuah perguruan tinggi. Kemajuan perguruan tinggi dapat dilihat dari jumlah mahasiswa yang melakukan pendaftaran pada setiap pembukaan tahun akademik baru[1]. Perguruan tinggi negeri dan swasta secara berkala pada setiap tahunnya melakukan kegiatan penerimaan mahasiswa baru. Upaya memaksimalkan kegiatan penerimaan mahasiswa baru butuh sebuah langkah-langkah khusus yang harus dilakukan perguruan tinggi untuk dapat mengetahui target dari mahasiswa baru secara detail[2]. Proses penerimaan mahasiswa baru merupakan aktivitas *value chain* perguruan tinggi[3]. Permasalahan yang terjadi dalam perekrutan mahasiswa adalah banyaknya jumlah calon mahasiswa yang tidak mendaftar untuk setiap tahunnya padahal peserta seleksi telah dinyatakan lulus. Hal ini menjadi masalah yang harus diselesaikan oleh pihak perguruan tinggi, yang berarti kuota mahasiswa baru untuk setiap program studi belum tercapai. Salah satu penyebab terjadinya penurunan jumlah mahasiswa baru pada tahun ini adalah kurang dilakukannya pengolahan data mahasiswa secara tepat berdasarkan data historis oleh pihak administrasi perguruan tinggi. Hal tersebut dapat mempengaruhi pengambilan keputusan dalam menentukan wilayah promosi yang tepat sasaran. Pengolahan data mahasiswa seharusnya dilakukan agar dapat menentukan wilayah promosi yang tepat sasaran sehingga tidak terjadi penurunan jumlah mahasiswa pada tahun berikutnya[4].

Akademi Manajemen Informatika dan Komputer HASS atau AMIK HASS merupakan perguruan tinggi swasta di Kota Bandung. Berbagai upaya strategi promosi telah dilakukan untuk menarik minat calon mahasiswa yaitu memasang spanduk, *broadcast* melalui media sosial maupun media cetak, penawaran beasiswa Yayasan dan kunjungan langsung ke sekolah. Kunjungan langsung ke sekolah bertujuan untuk mempromosikan keberadaan AMIK HASS Bandung. Penyebaran angket penelusuran potensi dan minat calon mahasiswa baru yang telah diisi oleh siswa menjadi referensi bagi Bagian Marketing untuk mendapatkan informasi calon mahasiswa yang akan mendaftar. Berdasarkan data yang diperoleh, penerimaan mahasiswa baru AMIK HASS mengalami penurunan jumlah pendaftar pada setiap tahun terutama masa pandemi Covid-19 saat ini. Permasalahan tersebut diperlukan upaya untuk mengolah data penerimaan mahasiswa baru yang ada, kemudian mengidentifikasi pola-pola dalam data tersebut sebagai informasi yang dapat memprediksi kemungkinan jumlah calon mahasiswa baru yang akan melakukan pendaftaran. Proses identifikasi pola data dapat dilakukan dengan menerapkan metode yang termasuk dalam ilmu data mining, salah satunya adalah klasifikasi[5].

Perguruan tinggi yang menjadi tempat penerapan *data mining* sangat memperhatikan data pendaftaran mahasiswa, mencari pola dan kemungkinan pengaruhnya terhadap keputusan mahasiswa untuk masuk perguruan tinggi. Prediksi, sebagai salah satu teknik data mining yang umum digunakan dalam literatur, dianggap sebagai metode praktis untuk manajemen perguruan tinggi dalam menghasilkan pengetahuan yang akan digunakan untuk pengambilan keputusan[6].

Tujuan penelitian ini diharapkan dapat memperoleh hasil dengan tingkat akurasi dan presisi yang lebih baik berdasarkan kinerja masing-masing algoritma klasifikasi. Sehingga dapat memberikan rekomendasi pengambilan keputusan bagi Bagian Marketing untuk menentukan arah strategi promosi di masa yang akan datang.

2 Tinjauan Literatur

Data mining sering digunakan dalam data berukuran besar untuk menemukan informasi penting dalam data tersebut. Sumber data dapat mencakup database, *data warehouse*, *database online* dan repositori informasi lainnya[7]. Penerapan *data mining* pada penelitian terdahulu telah banyak diterapkan pada perguruan tinggi untuk menunjang kegiatan penerimaan mahasiswa baru, salah satunya untuk melakukan prediksi dengan menggunakan algoritma klasifikasi. Prediksi peningkatan pendaftaran pada perguruan tinggi menggunakan *Neural Networks*, *Decision Trees*, *Bayes* dan *Logistic Regression* di Islamic Azad University, Iran. Hasil penelitian menunjukkan *Neural Network*, *Decision Tree*, *Bayes* dan *Logistic Regression* sangat efektif untuk melakukan prediksi terhadap calon pendaftar pada perguruan tinggi[8]. Kemudian, prediksi pendaftaran mahasiswa baru di Cebu Technological University-Barili Campus, Philippines. Teknik yang digunakan dalam prediksi yaitu model *Autoregressive Integrated Moving Average* atau ARIMA. *Data set* yang digunakan merupakan data masa lampau yaitu data pendaftaran mahasiswa baru mulai tahun akademik 2011-2012 sampai 2018-2019. Hasil penelitian menunjukkan tren peningkatan pendaftaran yang akan terjadi pada setiap tahunnya[9]. Prediksi peningkatan jumlah pendaftaran calon mahasiswa baru di Murang'a University

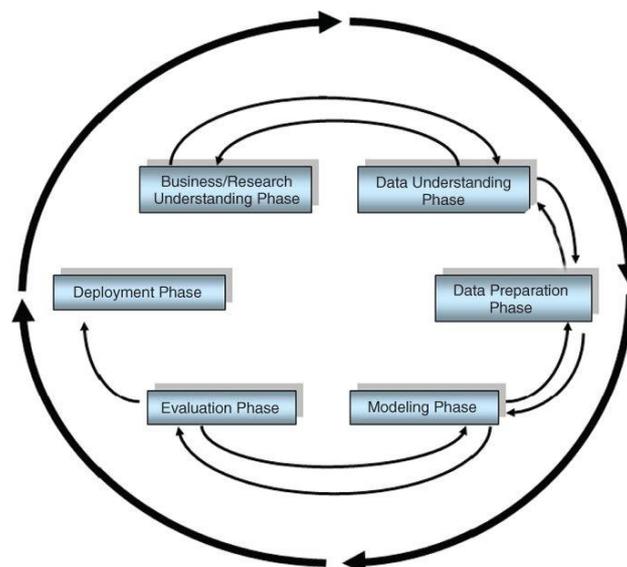
of Technology, Kenya. menggunakan *Ensemble Classifier*. Algoritma yang digunakan yaitu *J48*, *Naïve Bayes*, *Classification and Regression Tree* atau *CART*, dan *Bagging*. Hasil penelitian menunjukkan *J48* memiliki tingkat akurasi tertinggi sebesar 84% diikuti *Baging* sebesar 82%, *CART* sebesar 77% dan *Naïve Bayes* sebesar 72%[10]. Prediksi perilaku pendaftaran mahasiswa dan pola tingkat *drop out* pada institusi perguruan tinggi menggunakan *Machine Learning*. Penelitian tersebut menggunakan algoritma *Apriori Boosted Regression Tree*. Data set diperoleh dari *Abu Dhabi School of Management (ASDM)*, Uni Arab Emirates. Hasil penelitian menunjukkan tingkat akurasi pada tahun 2019 sebesar 86% dan tahun 2020 sebesar 78%. Penurunan tingkat akurasi pendaftaran mahasiswa pada tahun 2020 diakibatkan karena pandemi *Corona Virus-19*[11].

Penelitian terkait penerapan dan analisis perbandingan kinerja algoritma klasifikasi telah dilakukan di perguruan tinggi Indonesia untuk mengolah data penerimaan mahasiswa baru. Penerapan algoritma *K-Nearest Neighbor* untuk menentukan pola penerimaan mahasiswa baru di Polteknik Pos. *Data set* yang digunakan pada saat calon mahasiswa mendaftar dan melaksanakan ujian masuk. Algoritma *K-Nearest Neighbor* mendapatkan tingkat kepercayaan sebesar 67%-77%. Sehingga data penerimaan mahasiswa baru dapat dijadikan acuan untuk mengetahui profil mahasiswa baru yang akan melakukan Her Registrasi[12]. Kemudian, Implementasi *Naïve Bayes* dalam klasifikasi calon mahasiswa baru di STMIK Kharisma Makassar. Pengujian dilakukan dengan membagi *data set* sebanyak 80% untuk *data training* dan sebanyak 20% untuk *data testing*. Tingkat akurasi *Naïve Bayes* mencapai sebesar 73% dengan nilai *sensitivity* sebesar 46,93% dan *Precision* sebesar 95,83%. Hasil ini menunjukkan kinerja algoritma klasifikasi cukup baik pada *data set* calon mahasiswa baru[13]. Selain itu, *Naïve Bayes* digunakan dalam pengembangan sistem cerdas untuk prediksi daftar ulang mahasiswa baru di Universitas Pendidikan Ganesha. Hasil pengujian tingkat akurasi untuk keempat jalur pendaftaran memiliki rata-rata akurasi sebesar 71,65%, rata-rata presisi sebesar 75,46%, rata-rata *recall* sebesar 80,91%, rata-rata *specificity* sebesar 36,96%, dan rata-rata *f1 score* sebesar 80,94%. Sehingga hasil pengujian pada sistem tergolong cukup baik dalam memprediksi daftar ulang mahasiswa baru[14]. Penerapan algoritma ID3 untuk mendukung sosialisasi penerimaan mahasiswa baru di Universitas Islam Negeri Raden Fatah Palembang. Data yang digunakan dalam penelitian merupakan data wisuda tiga tahun terakhir yang berasal dari Pusat Teknologi Informasi dan Pangkalan Data (PUSTIPD). Evaluasi *data set* menggunakan *k-Fold Cross-Validation* dengan jumlah k-10 menghasilkan akurasi sebesar 67,26%. Sehingga algoritma klasifikasi ID3 menunjukkan hasil yang cukup baik[15]. Analisis dan penerapan algoritma *Support Vector Machine* dalam *data mining* untuk menunjang strategi promosi penerimaan mahasiswa baru di Universitas PGRI Semarang. Hasil pengujian diperoleh tingkat akurasi sebesar 73,6%[16]. Analisis kelayakan lokasi promosi penerimaan mahasiswa baru menggunakan algoritma *Naïve Bayes* dan *Decision Tree C4.5* di Sekolah Tinggi Manajemen dan Informatika (STMIK) Bani Saleh. Hasil pengujian menggunakan Rapidminer maka diperoleh nilai akurasi pada algoritma *Naïve Bayes* sebesar 91,43% dan *Decision Tree C4.5* sebesar 94,29%. Hasil pengukuran menunjukkan kedua algoritma cukup handal[17]. Analisis komparasi algoritma *data mining* untuk klasifikasi her registrasi calon mahasiswa di STMIK Widya Pratama. Algoritma yang digunakan dalam penelitian yaitu *Naïve Bayes*, *K-Nearest Neighbor* dan *Decision Tree C4.5*. Hasil pengujian tingkat akurasi algoritma *Decision Tree C4.5* merupakan yang terbaik yaitu 80,72% diikuti algoritma *K-Nearest Neighbor* sebesar 80,46% dan *Naïve Bayes* sebesar 74,49%. Ketiga algoritma klasifikasi dapat digunakan untuk mengklasifikasikan her registrasi calon mahasiswa[1]. Komparasi kinerja algoritma *C4.5* dan *Naive Bayes* untuk prediksi kegiatan penerimaan mahasiswa baru di Universitas STIKUBANK Semarang. Algoritma *C45* memiliki nilai akurasi lebih tinggi sebesar 88,74% dibandingkan algoritma *Naive Bayes* yang memiliki akurasi sebesar 87,24%[5].

Berdasarkan penelitian sebelumnya, penerapan algoritma klasifikasi cukup efektif dalam penerimaan mahasiswa baru. Pada Penelitian ini, *data set* yang akan digunakan berbeda dengan penelitian sebelumnya yaitu *data set* penelusuran potensi dan minat calon mahasiswa baru yang diperoleh dari Bagian Marketing. Algoritma klasifikasi yang digunakan yaitu algoritma *Naïve Bayes* (NB), *Decision Tree J48* (J48), *K-Nearest Neighbor* (K-NN), *Random Forest* (RF) dan *Support Vector Machine* (SVM). Penelitian ini menggunakan *Machine Learning Waikato Environment for Knowledge Analysis* atau WEKA[18].

3 Metode Penelitian

Metode yang digunakan dalam penelitian mengadopsi proses *data mining* CRISP-DM atau *Cross Industry Standard Process for Data Mining* yang terdiri dari enam fase yang dapat dilihat pada Gambar 1.



Gambar 1. Metode CRISP-DM [19][20]

3.1. Fase Pemahaman Bisnis (*Business Understanding Phase*)

Fase pemahaman bisnis *data mining* bertujuan untuk melakukan analisis perbandingan kinerja algoritma klasifikasi terhadap *data set* penelusuran potensi dan minat calon mahasiswa baru serta mengidentifikasi pola-pola yang terbentuk dari data masa lampau menggunakan algoritma NB, J48, K-NN, RF dan SVM.

3.2. Fase Pemahaman Data (*Data Understanding Phase*)

Fase pemahaman data terdiri dari beberapa proses diantaranya pengumpulan data awal, deskripsi data, eksplorasi data dan verifikasi kualitas data. Proses pengumpulan data bertujuan untuk mempelajari struktur data pada Bagian Marketing melalui penelusuran potensi dan minat calon mahasiswa baru, selanjutnya menentukan atribut yang akan digunakan dalam proses *data mining*. Proses deskripsi data bertujuan untuk menganalisis sumber data awal yang diperoleh dari *database* Bagian Marketing. Sehingga dapat menentukan atribut yang akan diproses dalam *data mining*. Proses eksplorasi data bertujuan untuk menentukan Sumber data yang digunakan dalam penelitian ini adalah data penelusuran potensi dan minat calon mahasiswa pada tahun 2013-2018 yang diperoleh dari Bagian Marketing AMIK HASS Bandung. Proses verifikasi kualitas data bertujuan untuk memastikan data yang digunakan pada proses *data mining* tidak ada kesalahan, yaitu memastikan sumber *record* terisi dengan data dan tidak terdapat *anomaly data*. Jika terjadi data kosong maka sesuai dengan ketentuan *record* tersebut akan dihapus. Kemudian memastikan tidak ada duplikasi data, jika terdapat duplikasi maka salah satu data akan dihapus.

3.3. Fase Pengolahan Data (*Data Preparation Phase*)

Fase pengolahan data terdapat beberapa proses yang dilakukan meliputi pemilihan data, pembersihan data dan transformasi data. Pada fase pengolahan data dilakukan pada *data set* penelusuran potensi dan minat calon mahasiswa baru sebanyak 6866 *record* terdiri dari 13 atribut. Proses pemilihan data ini dilakukan untuk memilih atribut yang akan digunakan pada pembentukan model algoritma klasifikasi. Proses pembersihan data dilakukan untuk proses *cleaning* dari data awal sebanyak 6866 *record* semuanya tidak utuh atau lengkap, dimana data yang tidak lengkap dibuang atau tidak digunakan. Sehingga yang tersisa sebanyak 5934 *record*. Pada proses transformasi data terdapat beberapa atribut yang dapat disederhanakan seperti atribut Alamat menjadi atribut Domisili dan atribut Tanggal Pendaftaran menjadi atribut Status. Kemudian dilakukan konversi nilai dari

<http://sistemasi.ftik.unisi.ac.id>

atribut Alamat terhadap atribut Domisili dengan menentukan wilayah tempat tinggal berdasarkan Kota atau Kabupaten. Konversi nilai atribut tanggal pendaftaran terhadap atribut Status pun dilakukan menjadi Daftar atau Tidak Daftar. Nilai atribut Peminatan awalnya berisi data jurusan yang diminati calon mahasiswa baru dikonversi menjadi beberapa kategori diantaranya Komputer, Non Komputer dan Kerja dalam hal ini tidak melanjutkan untuk kuliah. Tabel 1 merupakan atribut terpilih yang akan digunakan dalam *data set* pengujian untuk algoritma klasifikasi diantaranya algoritma klasifikasi NB, J48, K-NN, RF dan SVM.

Tabel 1. Atribut Data Set

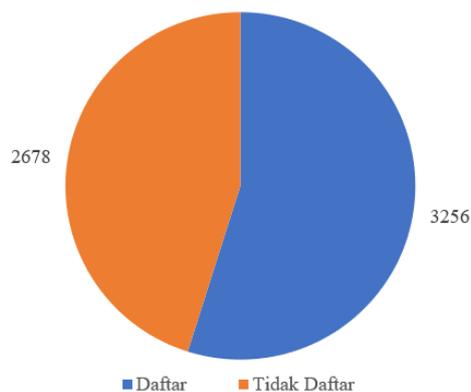
Atribut	Keterangan
Jenis Kelamin	Jenis kelamin calon mahasiswa baru
Domisili	Wilayah tempat tinggal calon mahasiswa baru
Asal Sekolah	Asal sekolah calon mahasiswa baru
Jurusan	Jurusan yang diambil calon mahasiswa baru pada saat sekolah terakhir
Pekerjaan Orang Tua	Pekerjaan orang tua calon mahasiswa baru
Peminatan	Jurusan yang akan dipilih calon mahasiswa baru setelah lulus dari sekolah
Minat Kuliah	Minat calon mahasiswa baru untuk kuliah di AMIK HASS Bandung
Status	Status pendaftaran calon mahasiswa baru

Setelah dilakukan proses pemilihan data, pembersihan data dan transformasi data. Tabel 2 merupakan sampel *record* yang terdapat dalam *data set* pengujian, atribut Status merupakan atribut target atau label dalam prediksi.

Tabel 2. Sampel Data Set Pengujian

Asal Sekolah	Jenis Kelamin	Domisili	Jurusan	Pekerjaan Orang Tua	Peminatan	Minat Kuliah	Status
SMA School	Lab P	Bandung	IPS	Karyawan Swasta	Non Komputer	Berminat	Tidak Daftar
SMAN Cicalengka	1 L	Kabupaten Bandung	IPA	Wiraswasta	Komputer	Berminat	Daftar
SMAN Serang Panjang	1 P	Bandung	IPS	Wiraswasta	Non Komputer	Berminat	Tidak Daftar
SMA School	Lab P	Bandung	IPS	Karyawan Swasta	Non Komputer	Berminat	Tidak Daftar
...
SMA Ciranjang	PGRI P	Cianjur	IPS	Wiraswasta	Komputer	Berminat	Daftar

Berdasarkan *data set* pengujian yang diperoleh, maka rekapitulasi atribut Status sebagai target atau *label* calon mahasiswa baru yang mendaftar dan tidak mendaftar ditunjukkan pada Gambar 2. Grafik menunjukkan calon mahasiswa yang mendaftar sebesar 54,87% atau 3.256 *record* dan yang tidak mendaftar sebesar 45,12% atau 2.678 *record*.



Gambar 2. Grafik Rekapitulasi Atribut Status sebagai Target Prediksi

3.4. Fase Pemodelan (*Modeling Phase*)

Penelitian berbentuk pengujian terhadap data penelusuran potensi dan minat calon mahasiswa baru memilih yang diambil dari Bagian Marketing yang dituangkan dalam bentuk tabel. Data tersebut akan dilakukan lima kali percobaan menggunakan algoritma klasifikasi NB, J48, K-NN, RF dan SVM.

Algoritma klasifikasi NB adalah pengklasifikasi probabilistik sederhana yang menghitung sekumpulan probabilitas dengan menghitung frekuensi dan kombinasi nilai dalam kumpulan data yang diberikan. Algoritma klasifikasi NB menggunakan teorema Bayes dan mengasumsikan bahwa semua variabel adalah independen dengan mempertimbangkan nilai variabel kelas[21]. Teorema Bayes adalah rumus matematika yang digunakan untuk menentukan probabilitas bersyarat, yang dinamai dari ahli matematika Inggris abad ke-18 Thomas Bayes dijelaskan pada rumus 1.

$$P(A|B) = \frac{P(A) P(B|A)}{P(B)} \quad (1)$$

$P(A|B)$ adalah peluang terjadinya peristiwa A pada saat peristiwa B terjadi, $P(A)$ adalah peluang terjadinya peristiwa A, $P(B|A)$ adalah peluang terjadinya peristiwa B pada saat peristiwa A terjadi, $P(B)$ adalah peluang terjadinya B.

Algoritma klasifikasi NB dapat bekerja dengan baik dari jumlah *data set* yang banyak maupun sedikit[13]. Han dan Kamber menjelaskan bahwa dalam penerapannya terhadap *database* yang besar, *Teorema Bayes* tetap mampu bekerja maksimal dan mempunyai keakuratan yang tinggi[7].

Algoritma klasifikasi J48 merupakan implementasi dari algoritma C4.5 yang dikembangkan oleh tim proyek WEKA[22]. *Decision Tree* merupakan salah satu metode data mining yang paling intuitif dan populer. Algoritma klasifikasi K-NN memiliki konsistensi yang kuat. Ketika jumlah data mendekati tak hingga, algoritma ini menjamin *error rate* yang tidak lebih dari dua kali *Bayes error rate*[12].

Algoritma klasifikasi RF adalah algoritma yang menggunakan metode pemisahan biner rekursif untuk mencapai node akhir dalam struktur pohon berdasarkan pada pohon klasifikasi dan regresi[23]. Algoritma klasifikasi RF merupakan salah satu model *Decision Tree* pada metode pembelajaran mesin yang populer digunakan untuk mengembangkan model prediksi di berbagai penelitian. Pada *Decision Tree* sering terjadi *overlap*, terutama jika kriteria dan kelas yang digunakan dalam jumlah yang besar. *Overlap* tersebut dapat diatasi dengan algoritma klasifikasi RF[24].

Algoritma klasifikasi SVM dapat dikategorikan baik dalam *forecasting*. Prinsip kerja algoritma klasifikasi SVM didasarkan atas prinsip *Structural Risk Minimization* (SRM) yang bertujuan untuk menghasilkan *Hyperplane* terbaik yang memisahkan *input space* pada dua buah *class*, dan secara sederhana prinsip dasar metode ini adalah *linier classifier* atau pengelompokan secara berarah yang kemudian dikembangkan sehingga mampu bekerja secara *problemon linier*. Prinsip ini sendiri merupakan kombinasi yang baik antara teori komputasi yang telah berkembang jauh sebelum adanya algoritma klasifikasi SVM. [16].

Skenario pengujian *data set* yang digunakan dalam penelitian ini sebanyak 5934 *record* dengan *mode test percentage split* yaitu 70% atau sebanyak 4154 *record* untuk *data training* dan 30% atau sebanyak 1780 *record* untuk *data testing*.

3.5. Fase Evaluasi (*Evaluation Phase*)

Fase evaluasi dilakukan terhadap keefektifan dan kualitas model yang digunakan. Selanjutnya didapatkan hasil prediksi untuk masing-masing algoritma klasifikasi. Hasil prediksi kemudian diuji tingkat akurasi dengan bantuan metode *Confusion Matrix*. Rumus *Confusion Matrix* dijelaskan pada rumus 2.

$$Akurasi = \frac{TP+TN}{TP+FP+TN+FN} \quad (2)$$

TP (*True Positive*) adalah data positif yang diprediksi benar, FP (*False Positive*) adalah data negatif namun diprediksi sebagai data positif, TN (*True Negative*) adalah data negatif yang diprediksi benar, dan FN (*False Negative*) adalah data positif namun diprediksi sebagai data negatif. Keempat nilai tersebut akan dihitung untuk menentukan tingkat akurasi di setiap algoritma klasifikasi. Sehingga dapat diperoleh nilai akurasi tertinggi.

Receiver Operating Characteristic (ROC) digunakan untuk menilai hasil prediksi. Kurva ROC dibagi dalam dua dimensi, dimana tingkat *True Positive* diplot pada sumbu Y dan tingkat *False Positive* diplot pada sumbu X. akan tetapi dalam melakukan representasi grafis untuk menentukan klasifikasi mana yang lebih baik maka digunakan metode yang melakukan penghitungan luas daerah dibawah kurva ROC yang disebut AUC (*Area Under ROC Curve*). Nilai AUC klasifikasi data mining dapat dibagi menjadi beberapa kelompok[25].

Tabel 3. Standar Nilai AUC

Nilai AUC	Kualitas AUC
0,90 – 1,00	Excellent Classification
0,80 – 0,90	Good Classification
0,70 – 0,80	Fair Classification
0,60 – 0,70	Poor Classification
0,50 – 0,60	Failuer Classification

3.6. Fase Penyebaran (*Deployment Phase*)

Fase penyebaran bertujuan untuk mempresentasikan hasil pengujian yang telah dibuat dalam bentuk laporan. Sehingga dapat memberikan informasi atau pengetahuan data set dari masing-masing algoritma. Hasil tersebut menjadi rekomendasi pihak perguruan tinggi untuk menentukan strategi promosi di masa yang akan datang.

4 Hasil dan Pembahasan

4.1. Hasil Pengujian Algoritma Klasifikasi NB

Hasil pengujian algoritma klasifikasi NB terhadap *data training* ditunjukkan pada Gambar 3. Pengujian dilakukan dengan metode yang sama terhadap algoritma klasifikasi J48, K-NN, RF dan SVM yaitu hasil output *Correctly Classified Instances*, *Mean Absolute Error*, *Root Mean Squared Error*, *Average Precision*, *Average Recall*, *Time Taken to Build Model*, *Time Taken to Test Model on Test Split*, *F-Measure* dan *Area Under ROC*.

```

=== Summary ===

Correctly Classified Instances      1553           87.2472 %
Incorrectly Classified Instances    227           12.7528 %
Kappa statistic                    0.7451
Mean absolute error                 0.1665
Root mean squared error             0.2953
Relative absolute error             33.6102 %
Root relative squared error         59.339 %
Total Number of Instances          1780

=== Detailed Accuracy By Class ===

          TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
0.836   0.083   0.924   0.836   0.878   0.749   0.946   0.961   DAFTAR
0.917   0.164   0.821   0.917   0.866   0.749   0.946   0.923   TIDAK DAFTAR
Weighted Avg.  0.872   0.120   0.878   0.872   0.873   0.749   0.946   0.944

=== Confusion Matrix ===

  a  b  <-- classified as
817 160 | a = DAFTAR
 67 736 | b = TIDAK DAFTAR
    
```

Gambar 3. Hasil Pengujian Algoritma Klasifikasi NB

Hasil pengujian menggunakan algoritma klasifikasi NB memiliki tingkat akurasi sebesar 87,24%. Rasio prediksi benar sebanyak 1553 record dari keseluruhan data testing yang diujikan. Tingkat kesalahan hasil prediksi dapat dilihat dari nilai *Root Mean Squared Error* yang mempunyai nilai 0,30 sedangkan tingkat rata-rata kesalahan antara nilai aktual dan nilai prediksi dapat dilihat dari *Mean Absolute Error* sebesar 0,17. *Precision* pada kelas Daftar sebesar 92,40% menunjukkan bahwa persentase calon mahasiswa baru yang benar mendaftarkan dari keseluruhan calon mahasiswa baru yang diprediksi mendaftarkan. Kelas Tidak Daftar memiliki *Precision* sebesar 82,10% menunjukkan bahwa persentase calon mahasiswa baru yang benar tidak mendaftarkan dari keseluruhan calon mahasiswa baru mahasiswa yang diprediksi tidak mendaftarkan.

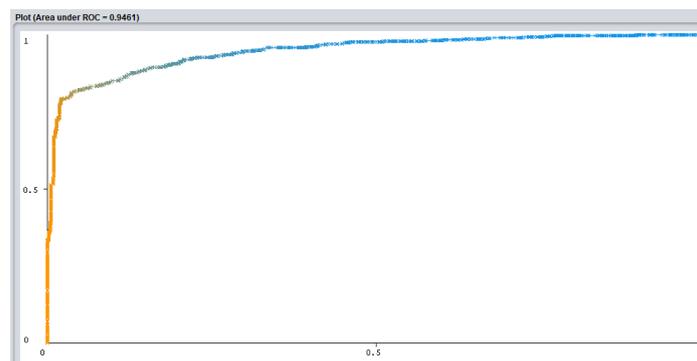
Recall pada kelas Daftar memiliki nilai sebesar 83,60% menunjukkan bahwa persentase calon mahasiswa baru yang diprediksi mendaftarkan dibandingkan keseluruhan calon mahasiswa baru yang sebenarnya mendaftarkan. *Recall* pada kelas Tidak Daftar memiliki nilai sebesar 91,70% menunjukkan bahwa persentase calon mahasiswa baru yang diprediksi tidak mendaftarkan dibandingkan keseluruhan calon mahasiswa baru yang sebenarnya tidak mendaftarkan.

Confusion Matrix algoritma klasifikasi NB baris pertama memiliki nilai 817 dan 160. Hal ini menunjukkan bahwa terdapat instance class Daftar dalam data testing diantaranya 817 record dinyatakan benar diprediksi sebagai kelas Daftar dan 160 record dinyatakan salah diprediksi sebagai kelas Tidak Daftar. Baris kedua terdapat nilai 67 dan 736 menunjukkan bahwa terdapat instance class Tidak Daftar dalam data testing diantaranya 67 record dinyatakan salah diprediksi sebagai kelas Daftar dan 736 record dinyatakan benar diprediksi kelas Tidak Daftar. Gambar 4 menunjukkan hasil prediksi pada data testing. Sedangkan Waktu yang dibutuhkan dalam test model selama 0,05 detik.

No.	1: ASAL SEKOLAH Nominal	2: JENIS KELAMIN KELAMIN	3: DOMISILI Nominal	4: JURUSAN Nominal	5: PEKERJAAN ORANG TUA ORANG TUA	6: PEMINATAN Nominal	7: MINAT KULIAH Nominal	8: prediction margin	9: predicted STATUS Nominal	10: STATUS Nominal
1	SMAN 1 TANJUNGPONORE	P	SUMEDANG	IPS	WIRASWASTA	NON KOMPUTER	BERMINAT	0.997943	DAFTAR	DAFTAR
2	SMA PGRI 1 SUBANG	L	SUBANG	IPS	PETANI	NON KOMPUTER	BERMINAT	0.989751	TIDAK DAFTAR	TIDAK DAFTAR
3	MAN CILILIN	L	BANDUNG BARAT	IPS	WIRASWASTA	KERJA	BERMINAT	0.274325	TIDAK DAFTAR	TIDAK DAFTAR
4	SMAN 1 JALAN CAGAK	P	SUBANG	IPA	WIRASWASTA	KOMPUTER	BERMINAT	0.987231	TIDAK DAFTAR	TIDAK DAFTAR
5	SMA PGRI PARAKAN MUNCANG	L	SUMEDANG	IPS	WIRASWASTA	KERJA	BERMINAT	0.998971	DAFTAR	DAFTAR
6	SMA PGRI CIRANJUR	L	CIRANJUR	IPS	WIRASWASTA	KOMPUTER	BERMINAT	-0.385752	DAFTAR	TIDAK DAFTAR
7	SMAN 1 CONGGEGANG	P	SUMEDANG	IPA	PNS	NON KOMPUTER	TIDAK BERMINAT	0.982789	DAFTAR	DAFTAR
8	MA PAJAGALAN	L	BANDUNG	LAIN-LAIN	WIRASWASTA	KERJA	TIDAK BERMINAT	0.986856	TIDAK DAFTAR	TIDAK DAFTAR
9	SMAN 1 TANJUNGPONORE	P	SUMEDANG	IPS	WIRASWASTA	NON KOMPUTER	TIDAK BERMINAT	0.995408	DAFTAR	DAFTAR
10	SMAN 1 CIWIDEY	P	KABUPATEN BANDUNG	IPA	WIRASWASTA	NON KOMPUTER	BERMINAT	0.771193	DAFTAR	DAFTAR
11	SMAN 1 KATAPANG	L	BANDUNG	IPA	KARYAWAN SWASTA	KOMPUTER	BERMINAT	-0.055939	TIDAK DAFTAR	DAFTAR
12	SMAN 1 CICALENGKA	L	KABUPATEN BANDUNG	IPA	IRT	NON KOMPUTER	TIDAK BERMINAT	0.929459	TIDAK DAFTAR	TIDAK DAFTAR
13	SMA PGRI PARAKAN MUNCANG	P	SUMEDANG	IPA	WIRASWASTA	KOMPUTER	BERMINAT	0.99925	DAFTAR	DAFTAR
14	SMA ANGKASA	P	BANDUNG	IPA	WIRASWASTA	NON KOMPUTER	TIDAK BERMINAT	0.947882	DAFTAR	DAFTAR
15	SMAN 4 CIMAH	L	CIMAH	IPS	WIRASWASTA	KOMPUTER	BERMINAT	0.251274	TIDAK DAFTAR	TIDAK DAFTAR
16	SMAN 1 JALAN CAGAK	P	SUBANG	IPA	WIRASWASTA	KOMPUTER	BERMINAT	0.987231	TIDAK DAFTAR	TIDAK DAFTAR
17	MAN CIPARAY	P	KABUPATEN BANDUNG	LAIN-LAIN	KARYAWAN SWASTA	KOMPUTER	BERMINAT	0.749693	TIDAK DAFTAR	TIDAK DAFTAR
18	MAN SUBANG	P	SUBANG	IPA	KARYAWAN SWASTA	NON KOMPUTER	TIDAK BERMINAT	0.940201	TIDAK DAFTAR	TIDAK DAFTAR
19	SMK AS-SHOFA	P	TASIKMALAYA	TKJ	WIRASWASTA	KOMPUTER	BERMINAT	0.973939	DAFTAR	DAFTAR
20	SMAN 1 RANCAKALONG	L	SUMEDANG	IPA	WIRASWASTA	KOMPUTER	BERMINAT	0.998594	DAFTAR	DAFTAR
21	SMKN 11 BANDUNG	P	BANDUNG	PEMASAR..	WIRASWASTA	KOMPUTER	BERMINAT	0.999879	DAFTAR	DAFTAR
22	SMA PGRI 1 SUBANG	L	SUBANG	IPS	PNS	NON KOMPUTER	BERMINAT	0.99186	TIDAK DAFTAR	TIDAK DAFTAR
23	SMKN 1 KATAPANG	L	KABUPATEN BANDUNG	ELEKTRO	WIRASWASTA	NON KOMPUTER	BERMINAT	0.982611	TIDAK DAFTAR	TIDAK DAFTAR
24	MAN CILILIN	P	BANDUNG BARAT	IPS	WIRASWASTA	KERJA	BERMINAT	-0.177054	TIDAK DAFTAR	DAFTAR
25	SMAN 1 CIPATAT	P	BANDUNG BARAT	IPS	WIRASWASTA	KERJA	BERMINAT	0.969113	DAFTAR	DAFTAR
26	MA PAJAGALAN	L	BANDUNG	IPS	WIRASWASTA	NON KOMPUTER	TIDAK BERMINAT	0.956146	TIDAK DAFTAR	TIDAK DAFTAR
27	SMAN 1 JALAN CAGAK	L	SUBANG	IPS	PNS	KOMPUTER	BERMINAT	0.99181	TIDAK DAFTAR	TIDAK DAFTAR
28	SMKN 11 BANDUNG	P	BANDUNG	RPL	KARYAWAN SWASTA	KOMPUTER	BERMINAT	0.99972	DAFTAR	DAFTAR
29	SMA LAB SCHOOL	P	BANDUNG	IPS	WIRASWASTA	NON KOMPUTER	TIDAK BERMINAT	0.938893	TIDAK DAFTAR	TIDAK DAFTAR
30	SMAN 1 CIPARAY	P	KABUPATEN BANDUNG	IPA	WIRASWASTA	KERJA	TIDAK BERMINAT	-0.269932	DAFTAR	TIDAK DAFTAR
31	SMKN 11 BANDUNG	P	BANDUNG	RPL	WIRASWASTA	KERJA	BERMINAT	0.999568	DAFTAR	DAFTAR
32	SMAN 1 PANCALENGAN	P	KABUPATEN BANDUNG	IPA	WIRASWASTA	NON KOMPUTER	BERMINAT	0.91165	DAFTAR	DAFTAR
33	SMAN 4 CIMAH	L	CIMAH	IPS	WIRASWASTA	NON KOMPUTER	TIDAK BERMINAT	0.763607	TIDAK DAFTAR	TIDAK DAFTAR

Gambar 4. Hasil Prediksi Algoritma Klasifikasi NB

AUC dihitung untuk mengukur perbedaan performansi. Gambar 5 merupakan kurva ROC yang menunjukkan akurasi dan membandingkan klasifikasi secara *visual* dengan *False Positive* sebagai garis *horizontal* dan *True Negative* sebagai garis *vertical*. Berdasarkan hasil pengujian menggunakan pengukuran algoritma klasifikasi NB didapatkan hasil AUC sebesar 0,95.



Gambar 5. AUC Algoritma Klasifikasi NB

4.2. Hasil Pengujian Algoritma Klasifikasi J48

Hasil pengujian algoritma klasifikasi J48 terhadap *data training* ditunjukkan pada Gambar 6. Pengujian dilakukan dengan metode yang sama terhadap algoritma klasifikasi NB, K-NN, RF dan SVM.

```

=== Summary ===

Correctly Classified Instances      1608           90.3371 %
Incorrectly Classified Instances    172            9.6629 %
Kappa statistic                    0.8056
Mean absolute error                0.1297
Root mean squared error            0.2631
Relative absolute error            26.1894 %
Root relative squared error        52.879 %
Total Number of Instances         1780

=== Detailed Accuracy By Class ===

              TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
              0.895   0.086   0.927     0.895   0.910     0.806   0.965   0.974   DAFTAR
              0.914   0.105   0.877     0.914   0.895     0.806   0.965   0.946   TIDAK DAFTAR
Weighted Avg.   0.903   0.095   0.904     0.903   0.904     0.806   0.965   0.961

=== Confusion Matrix ===

  a  b  <-- classified as
874 103 |  a = DAFTAR
 69 734 |  b = TIDAK DAFTAR
    
```

Gambar 6. Hasil Pengujian Algoritma Klasifikasi J48

Hasil pengujian menggunakan algoritma klasifikasi J48 memiliki tingkat akurasi sebesar 90,33%. Rasio prediksi benar sebanyak 1608 *record* dari keseluruhan *data testing* yang diujikan. Tingkat kesalahan hasil prediksi dapat dilihat dari nilai *Root Mean Squared Error* yang mempunyai nilai 0,26 sedangkan tingkat rata-rata kesalahan antara nilai aktual dan nilai prediksi dapat dilihat dari *Mean Absolute Error* memiliki nilai sebesar 0,13. *Precision* pada kelas Daftar 92,70% menunjukkan bahwa persentase calon mahasiswa baru yang benar mendaftar dari keseluruhan calon mahasiswa baru yang diprediksi mendaftar. Kelas Tidak Daftar memiliki nilai *Precision* sebesar 87,70% menunjukkan bahwa persentase calon mahasiswa baru yang benar tidak mendaftar dari keseluruhan calon mahasiswa baru mahasiswa yang diprediksi tidak mendaftar.

Recall pada kelas Daftar memiliki nilai sebesar 89,50% menunjukkan bahwa persentase calon mahasiswa baru yang diprediksi mendaftar dibandingkan keseluruhan calon mahasiswa baru yang sebenarnya mendaftar. *Recall* pada kelas Tidak daftar memiliki nilai sebesar 91,40% menunjukkan bahwa persentase calon mahasiswa baru yang diprediksi tidak mendaftar dibandingkan keseluruhan calon mahasiswa baru yang sebenarnya tidak mendaftar.

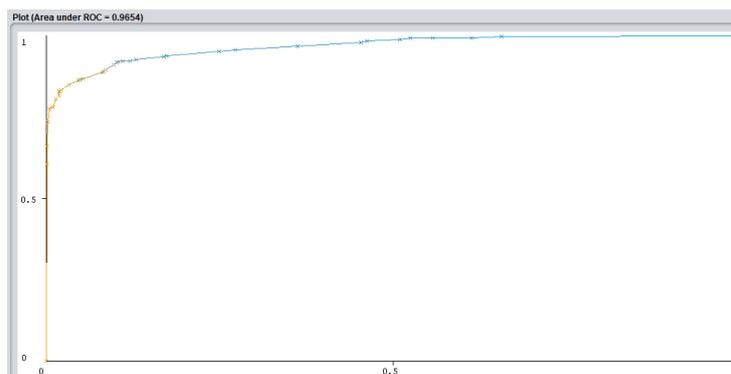
Confusion Matrix algoritma klasifikasi J48 pada baris pertama terdapat nilai 874 dan 103. Hal tersebut menunjukkan bahwa ada *instance class* Daftar dalam *data testing* diantaranya 874 *record*

dinyatakan benar diprediksi sebagai kelas Daftar dan 103 record dinyatakan salah diprediksi sebagai kelas Tidak Daftar. Baris kedua terdapat nilai 69 dan 734 menunjukkan bahwa ada instance class Tidak Daftar dalam data testing diantaranya 69 record dinyatakan salah diprediksi sebagai kelas Daftar dan 734 record dinyatakan benar diprediksi sebagai kelas Tidak Daftar. Gambar 7 merupakan hasil prediksi terhadap data testing sedangkan waktu yang dibutuhkan dalam test model selama 0.05 detik.

No.	1: ASAL SEKOLAH Nominal	2: JENIS KELAMIN Nominal	3: DOMISILI Nominal	4: JURUSAN Nominal	5: PEKERJAAN ORANG TUA Nominal	6: PEMINATAN Nominal	7: MINAT KULIAH Nominal	8: prediction margin	9: predicted STATUS Nominal	10: STATUS Nominal
1	SMAN 1 TANJUNGPURA	P	SUMEDANG	IPS	WIRASWASTA	NON KOMPUTER	BERMINAT	1.0	DAFTAR	DAFTAR
2	SMA PGRI 1 SUBANG	L	SUBANG	IPS	PETANI	NON KOMPUTER	BERMINAT	1.0	TIDAK DAFTAR	TIDAK DAFTAR
3	MAN CILILIN	L	BANDUNG BARAT	IPS	WIRASWASTA	KERJA	BERMINAT	0.75	TIDAK DAFTAR	TIDAK DAFTAR
4	SMAN 1 JALAN CAGAK	P	SUBANG	IPA	WIRASWASTA	KOMPUTER	BERMINAT	1.0	TIDAK DAFTAR	TIDAK DAFTAR
5	SMA PGRI PARAKAN MUNCANG	L	SUMEDANG	IPS	WIRASWASTA	KERJA	BERMINAT	1.0	DAFTAR	DAFTAR
6	SMA PGRI CIRANJUNG	L	CIAJUUR	IPS	WIRASWASTA	KOMPUTER	BERMINAT	-0.454545	DAFTAR	TIDAK DAFTAR
7	SMAN 1 CONGGEANG	P	SUMEDANG	IPA	PNS	NON KOMPUTER	TIDAK BERMINAT	1.0	DAFTAR	DAFTAR
8	MA PAJAGALAN	L	BANDUNG	LAIN-LAIN	WIRASWASTA	KERJA	TIDAK BERMINAT	1.0	TIDAK DAFTAR	TIDAK DAFTAR
9	SMAN 1 TANJUNGPURA	P	SUMEDANG	IPS	WIRASWASTA	NON KOMPUTER	TIDAK BERMINAT	1.0	DAFTAR	DAFTAR
10	SMAN 1 CIWIDEY	P	KABUPATEN BANDUNG	IPA	WIRASWASTA	NON KOMPUTER	BERMINAT	1.0	DAFTAR	DAFTAR
11	SMAN 1 KATAPANG	L	BANDUNG	IPA	KARYAWAN SWASTA	KOMPUTER	BERMINAT	1.0	DAFTAR	DAFTAR
12	SMAN 1 CICALENGKA	L	KABUPATEN BANDUNG	IPA	IRT	NON KOMPUTER	TIDAK BERMINAT	0.6	TIDAK DAFTAR	TIDAK DAFTAR
13	SMA PGRI PARAKAN MUNCANG	P	SUMEDANG	IPA	WIRASWASTA	KOMPUTER	BERMINAT	1.0	DAFTAR	DAFTAR
14	SMA ANGKASA	P	BANDUNG	IPA	WIRASWASTA	NON KOMPUTER	TIDAK BERMINAT	1.0	DAFTAR	DAFTAR
15	SMAN 4 CIMAH	L	CIMAH	IPS	WIRASWASTA	KOMPUTER	BERMINAT	-0.320388	DAFTAR	TIDAK DAFTAR
16	SMAN 1 JALAN CAGAK	P	SUBANG	IPA	WIRASWASTA	KOMPUTER	BERMINAT	1.0	TIDAK DAFTAR	TIDAK DAFTAR
17	MAN CIPARAY	P	KABUPATEN BANDUNG	LAIN-LAIN	KARYAWAN SWASTA	KOMPUTER	BERMINAT	1.0	TIDAK DAFTAR	TIDAK DAFTAR
18	MAN SUBANG	P	SUBANG	IPA	KARYAWAN SWASTA	NON KOMPUTER	TIDAK BERMINAT	0.649123	TIDAK DAFTAR	TIDAK DAFTAR
19	SMK AS-SHOFA	P	TASIKMALAYA	TKJ	WIRASWASTA	KOMPUTER	BERMINAT	1.0	DAFTAR	DAFTAR
20	SMAN 1 RANCAKALONG	L	SUMEDANG	IPA	WIRASWASTA	KOMPUTER	BERMINAT	1.0	DAFTAR	DAFTAR
21	SMKN 11 BANDUNG	P	BANDUNG	PEMASAR...	WIRASWASTA	KOMPUTER	BERMINAT	1.0	DAFTAR	DAFTAR
22	SMA PGRI 1 SUBANG	L	SUBANG	IPS	PNS	NON KOMPUTER	BERMINAT	1.0	TIDAK DAFTAR	TIDAK DAFTAR
23	SMKN 1 KATAPANG	L	KABUPATEN BANDUNG	ELEKTRO	WIRASWASTA	NON KOMPUTER	BERMINAT	0.891092	TIDAK DAFTAR	TIDAK DAFTAR
24	MAN CILILIN	P	BANDUNG BARAT	IPS	WIRASWASTA	KERJA	BERMINAT	-0.75	TIDAK DAFTAR	DAFTAR
25	SMAN 1 CIPATAT	P	BANDUNG BARAT	IPS	WIRASWASTA	KERJA	BERMINAT	1.0	DAFTAR	DAFTAR
26	MA PAJAGALAN	L	BANDUNG	IPS	WIRASWASTA	NON KOMPUTER	TIDAK BERMINAT	1.0	TIDAK DAFTAR	TIDAK DAFTAR
27	SMAN 1 JALAN CAGAK	L	SUBANG	IPS	PNS	KOMPUTER	BERMINAT	1.0	TIDAK DAFTAR	TIDAK DAFTAR
28	SMKN 11 BANDUNG	P	BANDUNG	RPL	KARYAWAN SWASTA	KOMPUTER	BERMINAT	1.0	DAFTAR	DAFTAR
29	SMA LAB SCHOOL	P	BANDUNG	IPS	WIRASWASTA	NON KOMPUTER	TIDAK BERMINAT	0.89781	TIDAK DAFTAR	TIDAK DAFTAR
30	SMAN 1 CIPARAY	P	KABUPATEN BANDUNG	IPA	WIRASWASTA	KERJA	TIDAK BERMINAT	-0.5	DAFTAR	TIDAK DAFTAR
31	SMKN 11 BANDUNG	P	BANDUNG	RPL	WIRASWASTA	KERJA	BERMINAT	1.0	DAFTAR	DAFTAR
32	SMAN 1 PANGALENGAN	P	KABUPATEN BANDUNG	IPA	WIRASWASTA	NON KOMPUTER	BERMINAT	0.948718	DAFTAR	DAFTAR
33	SMAN 4 CIMAH	L	CIMAH	IPS	WIRASWASTA	NON KOMPUTER	TIDAK BERMINAT	1.0	TIDAK DAFTAR	TIDAK DAFTAR
34	SMA LAB SCHOOL	L	BANDUNG	IPA	KARYAWAN SWASTA	KOMPUTER	BERMINAT	0.89781	TIDAK DAFTAR	TIDAK DAFTAR
35	SMAN 1 BOJONGSOANG	L	KABUPATEN BANDUNG	IPS	PNS	KOMPUTER	BERMINAT	1.0	TIDAK DAFTAR	TIDAK DAFTAR
36	MAN CIPARAY	P	KABUPATEN BANDUNG	IPA	LAIN-LAIN	KOMPUTER	TIDAK BERMINAT	-0.333333	DAFTAR	TIDAK DAFTAR
37	SMAN 1 BATULAJAR	L	BANDUNG BARAT	IPS	PNS	KOMPUTER	BERMINAT	1.0	TIDAK DAFTAR	TIDAK DAFTAR
38	SMK BAKARVA INTERNASIONAL	L	BANDUNG	OTOMOTIF	WIRASWASTA	NON KOMPUTER	BERMINAT	0.70026	DAFTAR	DAFTAR

Gambar 7. Hasil Prediksi Algoritma Klasifikasi J48

AUC dihitung untuk mengukur perbedaan performansi. Gambar 8 menunjukkan hasil pengujian AUC menggunakan pengukuran J48, nilai yang didapat sebesar 0,97.



Gambar 8. AUC Algoritma Klasifikasi J48

4.3. Hasil Pengujian Algoritma Klasifikasi K-NN

Hasil pengujian algoritma KNN terhadap data training ditunjukkan pada Gambar 9, pengujian dilakukan dengan metode yang sama pada algoritma klasifikasi NB, J48, RF dan SVM.

```

=== Summary ===

Correctly Classified Instances      1558           87.5281 %
Incorrectly Classified Instances     222           12.4719 %
Kappa statistic                     0.7481
Mean absolute error                  0.154
Root mean squared error              0.3125
Relative absolute error              31.0963 %
Root relative squared error          62.7964 %
Total Number of Instances           1780

=== Detailed Accuracy By Class ===

                TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
                0.887   0.139   0.886     0.887   0.887     0.748   0.934    0.943    DAFTAR
                0.861   0.113   0.863     0.861   0.862     0.748   0.934    0.903    TIDAK DAFTAR
Weighted Avg.   0.875   0.127   0.875     0.875   0.875     0.748   0.934    0.925

=== Confusion Matrix ===

  a  b  <-- classified as
867 110 |  a = DAFTAR
112 691 |  b = TIDAK DAFTAR
    
```

Gambar 9. Hasil Pengujian Algoritma Klasifikasi K-NN

Hasil pengujian menggunakan algoritma klasifikasi K-NN memiliki tingkat akurasi sebesar 87,52%. Rasio prediksi benar sebanyak 1558 *record* dari keseluruhan *data testing* yang diujikan. Tingkat kesalahan hasil prediksi dapat dilihat dari nilai *Root Mean Squared Error* yang mempunyai nilai 0,31 sedangkan tingkat rata-rata kesalahan antara nilai aktual dan nilai prediksi dapat dilihat dari *Mean Absolute Error* sebesar 0,15. *Precision* pada kelas Daftar sebesar 88,60% menunjukkan bahwa persentase calon mahasiswa baru yang benar mendaftar dari keseluruhan calon mahasiswa baru yang diprediksi mendaftar. Kelas Tidak Daftar memiliki *Precision* sebesar 86,30% menunjukkan bahwa persentase calon mahasiswa baru yang benar tidak mendaftar dari keseluruhan calon mahasiswa baru mahasiswa yang diprediksi tidak mendaftar.

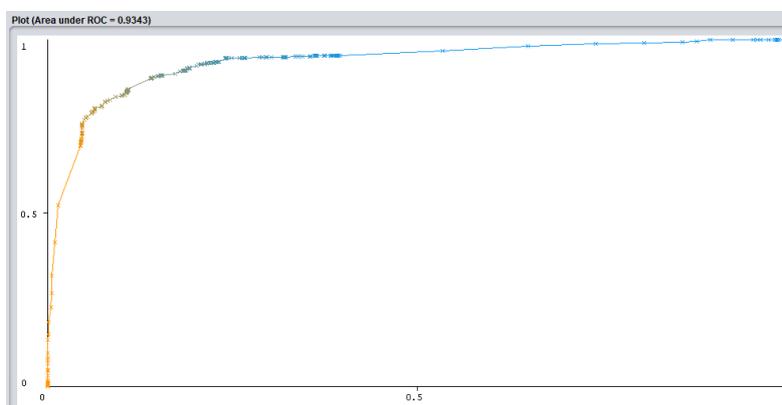
Recall pada kelas Daftar memiliki nilai sebesar 88,70% menunjukkan bahwa persentase calon mahasiswa baru yang diprediksi mendaftar dibandingkan keseluruhan calon mahasiswa baru yang sebenarnya mendaftar. Sedangkan *Recall* pada kelas Tidak daftar memiliki nilai sebesar 86,10% menunjukkan bahwa persentase calon mahasiswa baru yang diprediksi tidak mendaftar dibandingkan keseluruhan calon mahasiswa baru yang sebenarnya tidak mendaftar.

Confusion Matrix algoritma klasifikasi K-NN pada baris pertama terdapat nilai 867 dan 110. Hal ini menunjukkan bahwa ada *instance class* Daftar dalam *data testing* diantaranya 867 *record* dinyatakan benar diprediksi sebagai kelas Daftar dan 110 *record* dinyatakan salah diprediksi sebagai kelas Tidak Daftar. Pada Baris kedua terdapat nilai 112 dan 691 menunjukkan bahwa ada *instance class* Tidak Daftar dalam *data testing* diantaranya 112 *record* dinyatakan salah sebagai kelas Daftar dan 691 *record* dinyatakan benar diprediksi sebagai kelas Tidak Daftar. Gambar 10 merupakan hasil prediksi terhadap *data testing* sedangkan waktu yang dibutuhkan dalam test model selama 1,2 detik.

No.	1: ASAL SEKOLAH Nominal	2: JENIS KELAMIN Nominal	3: DOMISILI Nominal	4: JURUSAN Nominal	5: PEKERJAAN ORANG TUA Nominal	6: PEMINATAN Nominal	7: MINAT KULIAH Nominal	8: prediction margin Numeric	9: predicted STATUS Nominal	10: STATUS Nominal
1	SMAN 1 TANJUNGPURA	P	SUMEDANG	IPS	WIRASWASTA	NON KOMPUTER	BERMINAT	0.999952	DAFTAR	DAFTAR
2	SMA PGRI 1 SUBANG	L	SUBANG	IPS	PETANI	NON KOMPUTER	BERMINAT	0.999519	TIDAK DAFTAR	TIDAK DAFTAR
3	MAN CILILIN	L	BANDUNG BARAT	IPS	WIRASWASTA	KERJA	BERMINAT	0.714237	TIDAK DAFTAR	TIDAK DAFTAR
4	SMAN 1 JALAN CAGAK	P	SUBANG	IPA	WIRASWASTA	KOMPUTER	BERMINAT	0.999979	TIDAK DAFTAR	TIDAK DAFTAR
5	SMA PGRI PARAKAN MUNCANG	L	SUMEDANG	IPS	WIRASWASTA	KERJA	BERMINAT	0.999888	DAFTAR	DAFTAR
6	SMA PGRI CIRANJUNG	L	CIRANJUR	IPS	WIRASWASTA	KOMPUTER	BERMINAT	-0.142847	DAFTAR	TIDAK DAFTAR
7	SMAN 1 CONGGEANG	P	SUMEDANG	IPA	PNS	NON KOMPUTER	TIDAK BERMINAT	0.999759	DAFTAR	DAFTAR
8	MA PAJAGALAN	L	BANDUNG	LAIN-LAIN	WIRASWASTA	KERJA	TIDAK BERMINAT	0.999759	TIDAK DAFTAR	TIDAK DAFTAR
9	SMAN 1 TANJUNGPURA	P	SUMEDANG	IPS	WIRASWASTA	NON KOMPUTER	TIDAK BERMINAT	0.999904	DAFTAR	DAFTAR
10	SMAN 1 CIWIDEY	P	KABUPATEN BANDUNG	IPA	WIRASWASTA	NON KOMPUTER	BERMINAT	0.99992	DAFTAR	DAFTAR
11	SMAN 1 KATAPANG	L	BANDUNG	IPA	KARYAWAN SWASTA	KOMPUTER	BERMINAT	0.624981	DAFTAR	DAFTAR
12	SMAN 1 CICALENGKA	L	KABUPATEN BANDUNG	IPA	IRT	NON KOMPUTER	TIDAK BERMINAT	0.33332	TIDAK DAFTAR	TIDAK DAFTAR
13	SMA PGRI PARAKAN MUNCANG	P	SUMEDANG	IPA	WIRASWASTA	KOMPUTER	BERMINAT	0.99996	DAFTAR	DAFTAR
14	SMA ANGKASA	P	BANDUNG	IPA	WIRASWASTA	NON KOMPUTER	TIDAK BERMINAT	0.99984	DAFTAR	DAFTAR
15	SMAN 4 CIMAH	L	CIMAH	IPS	WIRASWASTA	KOMPUTER	BERMINAT	0.0	DAFTAR	TIDAK DAFTAR
16	SMAN 1 JALAN CAGAK	P	SUBANG	IPA	WIRASWASTA	KOMPUTER	BERMINAT	0.999979	TIDAK DAFTAR	TIDAK DAFTAR
17	MAN CIPARAY	P	KABUPATEN BANDUNG	LAIN-LAIN	KARYAWAN SWASTA	KOMPUTER	BERMINAT	0.99988	TIDAK DAFTAR	TIDAK DAFTAR
18	MAN SUBANG	P	SUBANG	IPA	KARYAWAN SWASTA	NON KOMPUTER	TIDAK BERMINAT	0.99996	TIDAK DAFTAR	TIDAK DAFTAR
19	SMK AS-SHOFA	P	TASIKMALAYA	TKJ	WIRASWASTA	KOMPUTER	BERMINAT	0.999852	DAFTAR	DAFTAR
20	SMAN 1 RANCAKALONG	L	SUMEDANG	IPA	WIRASWASTA	KOMPUTER	BERMINAT	0.99984	DAFTAR	DAFTAR
21	SMKN 11 BANDUNG	P	BANDUNG	PEMASAR...	WIRASWASTA	KOMPUTER	BERMINAT	0.99988	DAFTAR	DAFTAR
22	SMA PGRI 1 SUBANG	L	SUBANG	IPS	PNS	NON KOMPUTER	BERMINAT	0.555526	TIDAK DAFTAR	TIDAK DAFTAR
23	SMKN 1 KATAPANG	L	KABUPATEN BANDUNG	ELEKTRO	WIRASWASTA	NON KOMPUTER	BERMINAT	0.666613	TIDAK DAFTAR	TIDAK DAFTAR
24	MAN CILILIN	P	BANDUNG BARAT	IPS	WIRASWASTA	KERJA	BERMINAT	-0.99984	TIDAK DAFTAR	DAFTAR
25	SMAN 1 CIPATAT	P	BANDUNG BARAT	IPS	WIRASWASTA	KERJA	BERMINAT	0.99984	DAFTAR	DAFTAR
26	MA PAJAGALAN	L	BANDUNG	IPS	WIRASWASTA	NON KOMPUTER	TIDAK BERMINAT	0.99994	TIDAK DAFTAR	TIDAK DAFTAR
27	SMAN 1 JALAN CAGAK	L	SUBANG	IPS	PNS	KOMPUTER	BERMINAT	0.99992	TIDAK DAFTAR	TIDAK DAFTAR
28	SMKN 11 BANDUNG	P	BANDUNG	RPL	KARYAWAN SWASTA	KOMPUTER	BERMINAT	0.99997	DAFTAR	DAFTAR
29	SMA LAB SCHOOL	P	BANDUNG	IPS	WIRASWASTA	NON KOMPUTER	TIDAK BERMINAT	0.99984	TIDAK DAFTAR	TIDAK DAFTAR
30	SMAN 1 CIPARAY	P	KABUPATEN BANDUNG	IPA	WIRASWASTA	KERJA	TIDAK BERMINAT	-0.33328	DAFTAR	TIDAK DAFTAR
31	SMKN 11 BANDUNG	P	BANDUNG	RPL	WIRASWASTA	KERJA	BERMINAT	0.999519	DAFTAR	DAFTAR
32	SMAN 1 PANGALENGAN	P	KABUPATEN BANDUNG	IPA	WIRASWASTA	NON KOMPUTER	BERMINAT	0.99984	DAFTAR	DAFTAR
33	SMAN 4 CIMAH	L	CIMAH	IPS	WIRASWASTA	NON KOMPUTER	TIDAK BERMINAT	0.999904	TIDAK DAFTAR	TIDAK DAFTAR
34	SMA LAB SCHOOL	L	BANDUNG	IPA	KARYAWAN SWASTA	KOMPUTER	BERMINAT	0.0	DAFTAR	TIDAK DAFTAR
35	SMAN 1 BOJONGSOANG	L	KABUPATEN BANDUNG	IPS	PNS	KOMPUTER	BERMINAT	0.999759	TIDAK DAFTAR	TIDAK DAFTAR
36	MAN CIPARAY	P	KABUPATEN BANDUNG	IPA	LAIN-LAIN	KOMPUTER	TIDAK BERMINAT	-0.999519	DAFTAR	TIDAK DAFTAR
37	SMAN 1 BATUJAJAR	L	BANDUNG BARAT	IPS	PNS	KOMPUTER	BERMINAT	0.538442	TIDAK DAFTAR	TIDAK DAFTAR
38	SMA PGRI PARAKAN MUNCANG	L	BANDUNG	OTOMOTIF	WIRASWASTA	NON KOMPUTER	BERMINAT	0.99984	DAFTAR	DAFTAR

Gambar 10. Hasil Prediksi Algoritma Klasifikasi K-NN

AUC dihitung untuk mengukur perbedaan performansi. Gambar 11 merupakan hasil pengujian AUC menggunakan pengukuran algoritma klasifikasi K-NN didapatkan hasil AUC sebesar 0,93.



Gambar 11. AUC Algoritma Klasifikasi K-NN

4.4. Hasil Pengujian Algoritma Klasifikasi RF

Hasil pengujian algoritma RF terhadap *data training* ditunjukkan pada Gambar 12, pengujian dilakukan dengan metode yang sama pada algoritma klasifikasi NB, J48, K-NN dan SVM.

```

=== Summary ===

Correctly Classified Instances      1585           89.0449 %
Incorrectly Classified Instances    195           10.9551 %
Kappa statistic                    0.7793
Mean absolute error                 0.1366
Root mean squared error             0.2812
Relative absolute error             27.5831 %
Root relative squared error         56.5049 %
Total Number of Instances          1780

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
      0.890    0.110    0.908     0.890   0.899     0.779   0.955    0.964    DAFTAR
      0.890    0.110    0.870     0.890   0.880     0.779   0.955    0.935    TIDAK DAFTAR
Weighted Avg.   0.890    0.110    0.891     0.890   0.891     0.779   0.955    0.951

=== Confusion Matrix ===

  a  b  <-- classified as
870 107 |  a = DAFTAR
 88 715 |  b = TIDAK DAFTAR
    
```

Gambar 12. Hasil Pengujian Algoritma Klasifikasi RF

Hasil pengujian menggunakan algoritma klasifikasi RF memiliki tingkat akurasi sebesar 89,04%. Rasio prediksi benar sebesar 1585 *record* dari keseluruhan *data testing* yang diujikan. Tingkat kesalahan hasil prediksi dapat dilihat dari nilai *Root Mean Squared Error* yang mempunyai nilai 0,28 sedangkan tingkat rata-rata kesalahan antara nilai aktual dan nilai prediksi dapat dilihat dari *Mean Absolute Error* sebesar 0,14. *Precision* pada kelas Daftar sebesar 90,80% menunjukkan bahwa persentase calon mahasiswa baru yang benar mendaftar dari keseluruhan calon mahasiswa baru yang diprediksi mendaftar. Kelas Tidak Daftar memiliki *Precision* sebesar 87,00% menunjukkan bahwa persentase calon mahasiswa baru yang benar tidak mendaftar dari keseluruhan calon mahasiswa baru mahasiswa yang diprediksi tidak mendaftar.

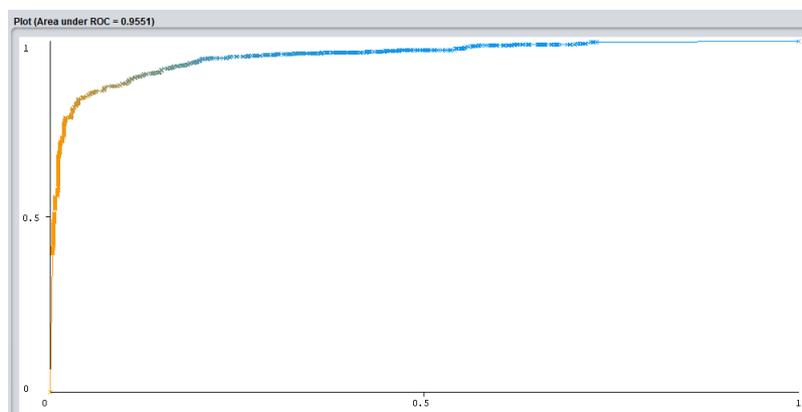
Recall pada kelas Daftar memiliki nilai sebesar 89,00% menunjukkan bahwa persentase calon mahasiswa baru yang diprediksi mendaftar dibandingkan keseluruhan calon mahasiswa baru yang sebenarnya mendaftar. *Recall* pada kelas Tidak daftar sebesar 89,00% menunjukkan bahwa persentase calon mahasiswa baru yang diprediksi tidak mendaftar dibandingkan keseluruhan calon mahasiswa baru yang sebenarnya tidak mendaftar.

Confusion Matrix algoritma klasifikasi RF pada baris pertama terdapat nilai 870 dan 107 menunjukkan bahwa ada *instance class* Daftar dalam *data testing* diantaranya 870 *record* dinyatakan benar diprediksi sebagai kelas Daftar dan 107 *record* dinyatakan salah diprediksi sebagai kelas Tidak Daftar. Pada Baris kedua terdapat nilai 88 dan 715 menunjukkan bahwa ada *instance class* Tidak Daftar dalam *data testing* diantaranya 88 *record* dinyatakan salah diprediksi sebagai kelas Daftar dan 715 *record* dinyatakan benar diprediksi sebagai kelas Tidak Daftar. Gambar 13 menunjukkan hasil prediksi terhadap *data testing* sedangkan waktu yang dibutuhkan dalam *test model* selama 0,14 detik.

No.	1: ASAL SEKOLAH	2: JENIS KELAMIN	3: DOMISILI	4: JURUSAN	5: PEKERJAAN ORANG TUA	6: PEMINATAN	7: MINAT KULIAH	8: prediction margin	9: predicted STATUS	10: STATUS
	Nominal		Nominal	Nominal	Nominal	Nominal	Nominal	Numeric	Nominal	Nominal
1	SMAN 1 TANJU...	P	SUMEDANG	IPS	WIRASWASTA	NON KOMPUTER	BERMINAT	1.0	DAFTAR	DAFTAR
2	SMA PGRI 1 SU...	L	SUBANG	IPS	PETANI	NON KOMPUTER	BERMINAT	0.96401	TIDAK DAFTAR	TIDAK DAFTAR
3	MAN CILILIN	L	BANDUNG BARAT	IPS	WIRASWASTA	KERJA	BERMINAT	0.704308	TIDAK DAFTAR	TIDAK DAFTAR
4	SMAN 1 JALAN ...	P	SUBANG	IPA	WIRASWASTA	KOMPUTER	BERMINAT	1.0	TIDAK DAFTAR	TIDAK DAFTAR
5	SMA PGRI PAR...	L	SUMEDANG	IPS	WIRASWASTA	KERJA	BERMINAT	1.0	DAFTAR	DAFTAR
6	SMA PGRI CIRA...	L	CIANJUR	IPS	WIRASWASTA	KOMPUTER	BERMINAT	-0.106649	DAFTAR	TIDAK DAFTAR
7	SMAN 1 CONG...	P	SUMEDANG	IPA	PNS	NON KOMPUTER	TIDAK BERMINAT	0.98	DAFTAR	DAFTAR
8	MA PAJAGALAN	L	BANDUNG	LAIN-LAIN	WIRASWASTA	KERJA	TIDAK BERMINAT	0.973303	TIDAK DAFTAR	TIDAK DAFTAR
9	SMAN 1 TANJU...	P	SUMEDANG	IPS	WIRASWASTA	NON KOMPUTER	TIDAK BERMINAT	1.0	DAFTAR	DAFTAR
10	SMAN 1 CIWIDEY	P	KABUPATEN BANDUNG	IPA	WIRASWASTA	NON KOMPUTER	BERMINAT	1.0	DAFTAR	DAFTAR
11	SMAN 1 KATAP...	L	BANDUNG	IPA	KARYAWAN SWASTA	KOMPUTER	BERMINAT	0.889389	DAFTAR	DAFTAR
12	SMAN 1 CICAL...	L	KABUPATEN BANDUNG	IPA	IRT	NON KOMPUTER	TIDAK BERMINAT	0.203445	TIDAK DAFTAR	TIDAK DAFTAR
13	SMA PGRI PAR...	P	SUMEDANG	IPA	WIRASWASTA	KOMPUTER	BERMINAT	1.0	DAFTAR	DAFTAR
14	SMA ANGKASA	P	BANDUNG	IPA	WIRASWASTA	NON KOMPUTER	TIDAK BERMINAT	1.0	DAFTAR	DAFTAR
15	SMAN 4 CIMAH...	L	CIMAH	IPS	WIRASWASTA	KOMPUTER	BERMINAT	-0.013602	DAFTAR	TIDAK DAFTAR
16	SMAN 1 JALAN ...	P	SUBANG	IPA	WIRASWASTA	KOMPUTER	BERMINAT	1.0	TIDAK DAFTAR	TIDAK DAFTAR
17	MAN CIPARAY	P	KABUPATEN BANDUNG	LAIN-LAIN	KARYAWAN SWASTA	KOMPUTER	BERMINAT	1.0	TIDAK DAFTAR	TIDAK DAFTAR
18	MAN SUBANG	P	SUBANG	IPA	KARYAWAN SWASTA	NON KOMPUTER	TIDAK BERMINAT	0.878278	TIDAK DAFTAR	TIDAK DAFTAR
19	SMK AS-SHOFA	P	TASIKMALAYA	TKJ	WIRASWASTA	KOMPUTER	BERMINAT	1.0	DAFTAR	DAFTAR
20	SMAN 1 RANCA...	L	SUMEDANG	IPA	WIRASWASTA	KOMPUTER	BERMINAT	1.0	DAFTAR	DAFTAR
21	SMKN 11 BAND...	P	BANDUNG	PEMASARAN	WIRASWASTA	KOMPUTER	BERMINAT	1.0	DAFTAR	DAFTAR
22	SMA PGRI 1 SU...	L	SUBANG	IPS	PNS	NON KOMPUTER	BERMINAT	0.950903	TIDAK DAFTAR	TIDAK DAFTAR
23	SMKN 1 KATAP...	L	KABUPATEN BANDUNG	ELEKTRO	WIRASWASTA	NON KOMPUTER	BERMINAT	0.636228	TIDAK DAFTAR	TIDAK DAFTAR
24	MAN CILILIN	P	BANDUNG BARAT	IPS	WIRASWASTA	KERJA	BERMINAT	-0.97	TIDAK DAFTAR	DAFTAR
25	SMAN 1 CIPAT...	P	BANDUNG BARAT	IPS	WIRASWASTA	KERJA	BERMINAT	1.0	DAFTAR	DAFTAR
26	MA PAJAGALAN	L	BANDUNG	IPS	WIRASWASTA	NON KOMPUTER	TIDAK BERMINAT	1.0	TIDAK DAFTAR	TIDAK DAFTAR
27	SMAN 1 JALAN ...	L	SUBANG	IPS	PNS	KOMPUTER	BERMINAT	1.0	DAFTAR	TIDAK DAFTAR
28	SMKN 11 BAND...	P	BANDUNG	RPL	KARYAWAN SWASTA	KOMPUTER	BERMINAT	1.0	DAFTAR	DAFTAR
29	SMA LAB SCHO...	P	BANDUNG	IPS	WIRASWASTA	NON KOMPUTER	TIDAK BERMINAT	0.993684	TIDAK DAFTAR	TIDAK DAFTAR
30	SMAN 1 CIPAR...	P	KABUPATEN BANDUNG	IPA	WIRASWASTA	KERJA	TIDAK BERMINAT	-0.474	DAFTAR	TIDAK DAFTAR
31	SMKN 11 BAND...	P	BANDUNG	RPL	WIRASWASTA	KERJA	BERMINAT	1.0	DAFTAR	DAFTAR
32	SMAN 1 PANGA...	P	KABUPATEN BANDUNG	IPA	WIRASWASTA	NON KOMPUTER	BERMINAT	0.993846	DAFTAR	DAFTAR
33	SMAN 4 CIMAH...	L	CIMAH	IPS	WIRASWASTA	NON KOMPUTER	TIDAK BERMINAT	0.99	TIDAK DAFTAR	TIDAK DAFTAR
34	SMA LAB SCHO...	L	BANDUNG	IPA	KARYAWAN SWASTA	KOMPUTER	BERMINAT	0.1195	TIDAK DAFTAR	TIDAK DAFTAR
35	SMAN 1 BOJON...	L	KABUPATEN BANDUNG	IPS	PNS	KOMPUTER	BERMINAT	0.901492	TIDAK DAFTAR	TIDAK DAFTAR
36	MAN CIPARAY	P	KABUPATEN BANDUNG	IPA	LAIN-LAIN	KOMPUTER	TIDAK BERMINAT	0.143022	TIDAK DAFTAR	TIDAK DAFTAR
37	SMAN 1 BATUJ...	L	BANDUNG BARAT	IPS	PNS	KOMPUTER	BERMINAT	0.971239	TIDAK DAFTAR	TIDAK DAFTAR
38	SMK PRAMAB...	L	BANDUNG	OTOMOTIF	WIRASWASTA	NON KOMPUTER	BERMINAT	0.883795	DAFTAR	DAFTAR

Gambar 13. Hasil Prediksi Algoritma Klasifikasi RF

AUC dihitung untuk mengukur perbedaan performansi. Gambar 14 menunjukkan hasil pengujian menggunakan pengukuran algoritma klasifikasi RF didapatkan hasil AUC sebesar 0,96.



Gambar 14. AUC Algoritma Klasifikasi RF

4.5. Hasil Pengujian Algoritma Klasifikasi SVM

Hasil pengujian algoritma klasifikasi SVM terhadap *data training* ditunjukkan pada Gambar 15, pengujian dilakukan dengan metode yang sama pada algoritma klasifikasi NB, J48, K-NN dan RF.

```

=== Summary ===

Correctly Classified Instances      1574           88.427 %
Incorrectly Classified Instances    206            11.573 %
Kappa statistic                     0.7685
Mean absolute error                  0.1157
Root mean squared error              0.3402
Relative absolute error              23.3681 %
Root relative squared error          68.3657 %
Total Number of Instances          1780

=== Detailed Accuracy By Class ===

                TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                0.852   0.076   0.932     0.852   0.890     0.772   0.888    0.875   DAFTAR
                0.924   0.148   0.837     0.924   0.878     0.772   0.888    0.807   TIDAK DAFTAR
Weighted Avg.   0.884   0.109   0.889     0.884   0.885     0.772   0.888    0.844

=== Confusion Matrix ===

  a  b  <-- classified as
832 145 |  a = DAFTAR
 61 742 |  b = TIDAK DAFTAR
    
```

Gambar 15. Hasil Pengujian Algoritma Klasifikasi SVM

Hasil pengujian algoritma klasifikasi SVM memiliki tingkat akurasi sebesar 88,43%. Rasio prediksi benar sebanyak 1585 *record* dari keseluruhan *data testing* yang diujikan. Tingkat kesalahan hasil prediksi dapat dilihat dari nilai *Root Mean Squared Error* yang mempunyai nilai 0,34 sedangkan tingkat rata-rata kesalahan antara nilai aktual dan nilai prediksi dapat dilihat dari *Mean Absolute Error* sebesar 0,12. *Precision* pada kelas Daftar sebesar 93,20% menunjukkan bahwa persentase calon mahasiswa baru yang benar mendaftar dari keseluruhan calon mahasiswa baru yang diprediksi mendaftar. Kelas Tidak Daftar memiliki *Precision* sebesar 83,70% menunjukkan bahwa persentase calon mahasiswa baru yang benar tidak mendaftar dari keseluruhan calon mahasiswa baru mahasiswa yang diprediksi tidak mendaftar.

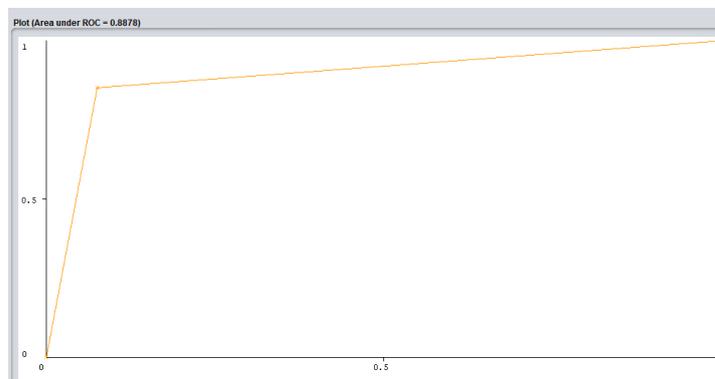
Recall pada kelas Daftar memiliki nilai sebesar 85,20% menunjukkan bahwa persentase calon mahasiswa baru yang diprediksi mendaftar dibandingkan keseluruhan calon mahasiswa baru yang sebenarnya mendaftar. Sedangkan *Recall* untuk kelas Tidak daftar sebesar 92,40% menunjukkan bahwa persentase calon mahasiswa baru yang diprediksi tidak mendaftar dibandingkan keseluruhan calon mahasiswa baru yang sebenarnya tidak mendaftar.

Confusion Matrix algoritma klasifikasi SVM pada baris pertama terdapat nilai 832 dan 145. Hal ini menunjukkan bahwa ada *instance class* Daftar dalam *data testing* diantaranya 832 *record* dinyatakan benar di prediksi sebagai kelas Daftar dan 145 *record* dinyatakan salah diprediksi sebagai kelas Tidak Daftar. Pada Baris kedua terdapat nilai 61 dan 742 menunjukkan bahwa ada *instance class* Tidak Daftar dalam *data testing* diantaranya 61 *record* dinyatakan salah diprediksi sebagai kelas Daftar dan 742 *record* dinyatakan benar diprediksi sebagai kelas Tidak Daftar. Hasil prediksi terhadap *data testing* pada algoritma klasifikasi SVM ditunjukkan pada Gambar 16, sedangkan waktu yang dibutuhkan dalam *test model* selama 0,06 detik.

No.	1: ASAL SEKOLAH Nominal	2: JENIS KELAMIN	3: DOMISILI Nominal	4: JURUSAN Nominal	5: PEKERJAAN ORANG TUA	6: PEMINATAN Nominal	7: MINAT KULIAH Nominal	8: prediction margin	9: predicted STATUS Nominal	10: STATUS Nominal
1	SMAN 1 TANJUNGPURA	P	SUMEDANG	IPS	WIRASWASTA	NON KOMPUTER	BERMINAT	1.0	DAFTAR	DAFTAR
2	SMA PGRI 1 SUBANG	L	SUBANG	IPS	PETANI	NON KOMPUTER	BERMINAT	1.0	TIDAK DAFTAR	TIDAK DAFTAR
3	MAN CILILIN	L	BANDUNG BARAT	IPS	WIRASWASTA	KERJA	BERMINAT	1.0	TIDAK DAFTAR	TIDAK DAFTAR
4	SMAN 1 JALAN CAGAK	P	SUBANG	IPA	WIRASWASTA	KOMPUTER	BERMINAT	1.0	TIDAK DAFTAR	TIDAK DAFTAR
5	SMA PGRI PARAKAN MUNCANG	L	SUMEDANG	IPS	WIRASWASTA	KERJA	BERMINAT	1.0	DAFTAR	DAFTAR
6	SMA PGRI CIRANJUNG	L	CIANJUR	IPS	WIRASWASTA	KOMPUTER	BERMINAT	-1.0	DAFTAR	TIDAK DAFTAR
7	SMAN 1 CONGGEGANG	P	SUMEDANG	IPA	PNS	NON KOMPUTER	TIDAK BERMINAT	1.0	DAFTAR	DAFTAR
8	MA PAJAGALAN	L	BANDUNG LAIN-LAIN	WIRASWASTA	KERJA	TIDAK BERMINAT	1.0	TIDAK DAFTAR	TIDAK DAFTAR	
9	SMAN 1 TANJUNGPURA	P	SUMEDANG	IPS	WIRASWASTA	NON KOMPUTER	TIDAK BERMINAT	1.0	DAFTAR	DAFTAR
10	SMAN 1 CIWIDEY	P	KABUPATEN BANDUNG	IPA	WIRASWASTA	NON KOMPUTER	BERMINAT	1.0	DAFTAR	DAFTAR
11	SMAN 1 KATAPANG	L	BANDUNG	IPA	KARYAWAN SWA...	KOMPUTER	BERMINAT	1.0	DAFTAR	DAFTAR
12	SMAN 1 CICALENGKA	L	KABUPATEN BANDUNG	IPA	IRT	NON KOMPUTER	TIDAK BERMINAT	1.0	TIDAK DAFTAR	TIDAK DAFTAR
13	SMA PGRI PARAKAN MUNCANG	P	SUMEDANG	IPA	WIRASWASTA	KOMPUTER	BERMINAT	1.0	DAFTAR	DAFTAR
14	SMA ANGKASA	P	BANDUNG	IPA	WIRASWASTA	NON KOMPUTER	TIDAK BERMINAT	1.0	DAFTAR	DAFTAR
15	SMAN 4 CIMAH	L	CIMAH	IPS	WIRASWASTA	KOMPUTER	BERMINAT	1.0	TIDAK DAFTAR	TIDAK DAFTAR
16	SMAN 1 JALAN CAGAK	P	SUBANG	IPA	WIRASWASTA	KOMPUTER	BERMINAT	1.0	TIDAK DAFTAR	TIDAK DAFTAR
17	MAN CIPARAY	P	KABUPATEN BANDUNG	LAIN-LAIN	KARYAWAN SWA...	KOMPUTER	BERMINAT	1.0	TIDAK DAFTAR	TIDAK DAFTAR
18	MAN SUBANG	P	SUBANG	IPA	KARYAWAN SWA...	NON KOMPUTER	TIDAK BERMINAT	1.0	TIDAK DAFTAR	TIDAK DAFTAR
19	SMK AS-SHOFA	P	TASKMALAYA	TKJ	WIRASWASTA	KOMPUTER	BERMINAT	1.0	DAFTAR	DAFTAR
20	SMAN 1 RANCAKALONG	L	SUMEDANG	IPA	WIRASWASTA	KOMPUTER	BERMINAT	1.0	DAFTAR	DAFTAR
21	SMKN 11 BANDUNG	P	BANDUNG	PEMASAR...	WIRASWASTA	KOMPUTER	BERMINAT	1.0	DAFTAR	DAFTAR
22	SMA PGRI 1 SUBANG	L	SUBANG	IPS	PNS	NON KOMPUTER	BERMINAT	1.0	TIDAK DAFTAR	TIDAK DAFTAR
23	SMAN 1 KATAPANG	L	KABUPATEN BANDUNG	ELEKTRO	WIRASWASTA	NON KOMPUTER	BERMINAT	1.0	TIDAK DAFTAR	TIDAK DAFTAR
24	MAN CILILIN	P	BANDUNG BARAT	IPS	WIRASWASTA	KERJA	BERMINAT	-1.0	TIDAK DAFTAR	DAFTAR
25	SMAN 1 CIPATAT	P	BANDUNG BARAT	IPS	WIRASWASTA	KERJA	BERMINAT	1.0	DAFTAR	DAFTAR
26	MA PAJAGALAN	L	BANDUNG	IPS	WIRASWASTA	NON KOMPUTER	TIDAK BERMINAT	1.0	TIDAK DAFTAR	TIDAK DAFTAR
27	SMAN 1 JALAN CAGAK	L	SUBANG	IPS	PNS	KOMPUTER	BERMINAT	1.0	TIDAK DAFTAR	TIDAK DAFTAR
28	SMKN 11 BANDUNG	P	BANDUNG	RPL	KARYAWAN SWA...	KOMPUTER	BERMINAT	1.0	DAFTAR	DAFTAR
29	SMA LAB SCHOOL	P	BANDUNG	IPS	WIRASWASTA	NON KOMPUTER	TIDAK BERMINAT	1.0	TIDAK DAFTAR	TIDAK DAFTAR
30	SMAN 1 CIPARAY	P	KABUPATEN BANDUNG	IPA	WIRASWASTA	KERJA	TIDAK BERMINAT	-1.0	DAFTAR	TIDAK DAFTAR
31	SMKN 11 BANDUNG	P	BANDUNG	RPL	WIRASWASTA	KERJA	BERMINAT	1.0	DAFTAR	DAFTAR
32	SMAN 1 PANGALENGAN	P	KABUPATEN BANDUNG	IPA	WIRASWASTA	NON KOMPUTER	BERMINAT	1.0	DAFTAR	DAFTAR
33	SMAN 4 CIMAH	L	CIMAH	IPS	WIRASWASTA	NON KOMPUTER	TIDAK BERMINAT	1.0	TIDAK DAFTAR	TIDAK DAFTAR
34	SMA LAB SCHOOL	L	BANDUNG	IPA	KARYAWAN SWA...	KOMPUTER	BERMINAT	1.0	TIDAK DAFTAR	TIDAK DAFTAR
35	SMAN 1 BOJONGSOANG	L	KABUPATEN BANDUNG	IPS	PNS	KOMPUTER	BERMINAT	1.0	TIDAK DAFTAR	TIDAK DAFTAR
36	MAN CIPARAY	P	KABUPATEN BANDUNG	IPA	LAIN-LAIN	KOMPUTER	TIDAK BERMINAT	1.0	TIDAK DAFTAR	TIDAK DAFTAR
37	SMAN 1 BATUJAJAR	L	BANDUNG BARAT	IPS	PNS	KOMPUTER	BERMINAT	1.0	TIDAK DAFTAR	TIDAK DAFTAR
38	SMK PRABUMIHAN INTERNASIONAL	L	BANDUNG	OTOMOTIF	WIRASWASTA	NON KOMPUTER	BERMINAT	1.0	DAFTAR	DAFTAR

Gambar 16. Hasil Prediksi Algoritma Klasifikasi SVM

AUC dihitung untuk mengukur perbedaan performansi. Gambar 17 merupakan hasil pengujian menggunakan pengukuran algoritma klasifikasi SVM didapatkan hasil AUC sebesar 0,89.



Gambar 17. AUC Algoritma Klasifikasi SVM

4.6. Hasil Perbandingan Algoritma Klasifikasi

Perbandingan kinerja algoritma klasifikasi pada penelitian ini berdasarkan hasil output pada kelima algoritma klasifikasi. Tabel 4 menunjukkan perbandingan kinerja kelima algoritma klasifikasi pada data set penelusuran potensi dan minat calon mahasiswa baru.

Tabel 4. Perbandingan Kinerja *Correctly Classified Instances, Mean Absolute Squared Error, Root Mean Squared Error dan Average Precision*

Algoritma Klasifikasi	Correctly Classified Instances (%)	Mean Absolute Error	Root Mean Squared Error	Average Precision (%)
NB	87,25	0,17	0,30	87,80
J48	90,34	0,13	0,26	90,40
K-NN	87,53	0,15	0,31	87,50
RF	89,04	0,14	0,28	89,10
SVM	88,43	0,12	0,34	88,90

Akurasi klasifikasi secara umum dapat ditentukan berdasarkan *Correctly Classified Instances*. Berdasarkan Tabel 4 Nilai akurasi prediksi tertinggi adalah algoritma klasifikasi J48 dengan nilai akurasi sebesar 90,34% diikuti RF sebesar 89,04%, SVM sebesar 88,43%, K-NN sebesar 87,53% dan algoritma klasifikasi NB memiliki akurasi terendah dari keempat klasifikasi yaitu sebesar 87,23%. Hal ini menunjukkan bahwa J48 merupakan algoritma klasifikasi yang memiliki akurasi lebih baik terhadap *data set* dibandingkan algoritma klasifikasi NB, K-NN, RF dan SVM.

Mean Absolute Error merupakan perbedaan rata-rata antara nilai absolut dan nilai prediksi. *Mean Absolute Error* pada kelima algoritma. Algoritma klasifikasi SVM memiliki nilai *Mean Absolute Error* paling rendah sebesar 0,12 diikuti J48 sebesar 0,13, RF sebesar 0,14, K-NN sebesar 0,15 dan NB sebesar 0,17. Semakin mendekati nilai nol pada *Mean Absolute Error* maka model klasifikasi semakin baik.

Root Mean Squared Error merupakan tingkat kesalahan hasil prediksi. Algoritma klasifikasi J48 memiliki nilai terendah yaitu 0,26 diikuti RF sebesar 0,28, NB sebesar 0,30, K-NN sebesar 0,31, dan SVM 0,34. Semakin tinggi nilai akurasi dan semakin rendah nilai tingkat kesalahan prediksi maka performansi model klasifikasi semakin baik.

Average precision merupakan rasio rata-rata dari hasil positif yang benar terhadap hasil prediksi yang positif. Perbandingan *Average Precision* pada kelima algoritma klasifikasi terlihat bahwa *Average Precision* tertinggi pada algoritma klasifikasi J48 sebesar 90,40% diikuti NB sebesar 87,80%, K-NN sebesar 87,50%, RF sebesar 89,10% dan SVM sebesar 88,90%. Semakin tinggi nilai *Average Precision* maka model klasifikasi semakin baik.

Tabel 5. Perbandingan Kinerja *Average Recall*, *Average F-Measure*, *AUROC*, *Time Taken to Test Model on Test Split* dan *Time Taken to Build Model*

Algoritma Klasifikasi	Average Recall (%)	Average F-Measure (%)	AUROC	Time Taken to Test Model on Test Split (second)	Time Taken to Build Model (second)
NB	87,20	87,30	0,95	0,03	0,02
J48	90,30	90,40	0,97	0,05	0,09
K-NN	87,50	87,50	0,93	1,20	0,00
RF	89,00	89,10	0,96	0,14	1,34
SVM	88,40	88,50	0,89	0,06	19,61

Average Recall merupakan rasio prediksi positif yang benar terhadap data keseluruhan yang benar-benar positif. Hal ini menunjukkan bahwa calon mahasiswa yang diprediksi mendaftar dibandingkan keseluruhan calon mahasiswa yang sebenarnya mendaftar. Perbandingan *Average Recall* pada kelima algoritma klasifikasi terlihat *Average Recall* tertinggi pada algoritma klasifikasi J48 sebesar 90,30% diikuti K-NN sebesar 87,50%, NB sebesar 87,20%, RF sebesar 89,00% dan SVM sebesar 88,40%. Semakin tinggi nilai *Average Recall* maka model klasifikasi semakin baik.

Time Taken to Build Model merupakan waktu yang dibutuhkan algoritma untuk membentuk model dari *data training* dalam satuan detik. Perbandingan *Time Taken to Build Model* pada kelima algoritma klasifikasi, terlihat waktu yang dibutuhkan algoritma klasifikasi K-NN paling cepat yaitu 0,00 detik diikuti NB selama 0,02 detik, J48 selama 0,09 detik, RF selama 1,34 detik, dan SVM selama 19,61 detik. Semakin cepat waktu yang dibutuhkan maka model semakin baik.

Time Taken to Test Model on Test Split merupakan waktu yang dibutuhkan model dalam melakukan pengujian *data testing* dalam satuan detik. Perbandingan *Time Taken to Test Model on Test Split* pada kelima algoritma klasifikasi, terlihat waktu yang dibutuhkan dalam pengujian *data testing*, algoritma klasifikasi NB paling cepat selama 0,03 detik diikuti J48 selama 0,05 detik, K-NN selama 1,20 detik, RF selama 0,14 detik dan SVM selama 0,06 detik. Semakin cepat waktu yang dibutuhkan maka model semakin baik.

F-Measure merupakan salah satu perhitungan evaluasi yang menggabungkan *recall* dan *precision*. Perbandingan hasil pengukuran *F-Measure* pada kelima algoritma klasifikasi, terlihat nilai *F-Measure* tertinggi pada algoritma klasifikasi J48 sebesar 90,40% diikuti K-NN sebesar 87,50%, NB sebesar 87,30%, RF sebesar 89,10% dan *Support Vector* sebesar 88,50%. Semakin tinggi hasil pengukuran *F-Measure* maka model klasifikasi semakin baik.

AUC merupakan pengujian untuk mengukur perbedaan performansi model. Perbandingan AUC pada kelima algoritma klasifikasi, terlihat pengujian performansi model tertinggi terdapat pada algoritma klasifikasi J48 sebesar 0,97 diikuti NB sebesar 0,95, K-NN sebesar 0,93, RF sebesar 0,96, SVM sebesar 0,89. Hasil pengujian AUC algoritma klasifikasi J48, NB, K-NN dan RF termasuk dalam kategori *Excellent Classification* sedangkan algoritma klasifikasi SVM termasuk dalam kategori *Good Classification* [25]. Semakin tinggi nilai AUC maka performansi model klasifikasi semakin baik.

5 Kesimpulan

Penelitian ini dilakukan untuk menganalisa perbandingan kinerja algoritma klasifikasi *data mining* pada *data set* penelusuran potensi dan minat calon mahasiswa baru. Berdasarkan hasil pengujian terhadap kelima algoritma klasifikasi yaitu NB, J48, K-NN, RF dan SVM. Nilai akurasi tertinggi diperoleh algoritma klasifikasi J48 sebesar 90,34% terdiri dari 1608 *record* yang terklarifikasi benar dari 1780 *data testing*, selain itu algoritma klasifikasi J48 memiliki tingkat kesalahan prediksi yang paling rendah sebesar 0,26. Meskipun demikian, perbedaan tingkat akurasi dengan keempat algoritma lainnya tidak menunjukkan perbedaan yang cukup signifikan. RF memiliki nilai akurasi sebesar 89,04% terdiri dari 1585 *record* yang terklarifikasi benar dari 1780 *data testing*. Algoritma klasifikasi SVM memiliki nilai akurasi sebesar 88,43% terdiri dari 1574 *record* yang terklarifikasi benar dari 1780 *data testing*. K-NN memiliki nilai akurasi sebesar 87,53% terdiri dari 1558 *record* yang terklarifikasi benar dari 1780 *data testing*. NB memiliki akurasi lebih rendah dari keempat algoritma klasifikasi sebesar 87,25% terdiri dari 1553 *record* yang terklarifikasi benar dari 1780 *data testing*.

Perbandingan hasil pengujian AUC, algoritma klasifikasi J48 memiliki nilai performansi tertinggi sebesar 0,97. Meskipun demikian, hasil pengujian AUC pada algoritma klasifikasi NB, K-NN dan RF tidak menunjukkan perbedaan yang cukup signifikan, sehingga keempat algoritma klasifikasi tersebut termasuk dalam kategori *Excellent Classification*. Berbeda pada algoritma klasifikasi SVM yang memiliki nilai AUC terendah dari keempat klasifikasi dan termasuk dalam kategori *Good Classification*.

Hasil penelitian yang telah dilakukan dapat disimpulkan bahwa algoritma klasifikasi J48 merupakan algoritma terbaik pada pengujian *data set* penelusuran potensi dan minat calon mahasiswa baru. Algoritma klasifikasi J48 memiliki aturan eksplisit untuk klasifikasi dan penanganan data heterogen yang tepat, berdasarkan hasil pengujian aturan yang dimiliki pada algoritma tersebut sebanyak 255 *rule*. Hasil penelitian sebelumnya dikemukakan oleh peneliti[10]. Pada penelitiannya dikemukakan bahwa algoritma klasifikasi J48 memiliki tingkat akurasi terbaik, meskipun demikian penelitian ini menggunakan objek, *data set*, dan atribut yang berbeda yaitu asal sekolah, jenis kelamin, domisili, jurusan, pekerjaan orang tua, peminatan, minat kuliah dan status.

Kontribusi penelitian ini, memberikan peluang kepada peneliti lain, untuk melakukan penelitian menggunakan algoritma klasifikasi dalam kegiatan penerimaan mahasiswa baru. Sehingga pihak perguruan tinggi dapat menerapkan berbagai strategi promosi untuk keberlangsungan kegiatan pembelajaran di masa pandemi Covid-19.

Ucapan Terima Kasih

Peneliti mengucapkan terima kasih kepada Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi yang telah memberikan kesempatan untuk melaksanakan penelitian dengan skema Penelitian Dasar Pemula tahun pelaksanaan 2021 serta Lembaga Penelitian dan Pengabdian Masyarakat (LPPM) Universitas Informatika dan Bisnis Indonesia yang telah mendorong dan memfasilitasi dalam setiap pelaksanaan penelitian.

Referensi

- [1] D. Aribowo and A. E. H. Setiadi, "Analisa Komparasi Algoritma Data Mining untuk Klasifikasi Heregistrasi Calon Mahasiswa STMIK Widya Pratama," *IC-Tech*, vol. 13, no. 2, pp. 1–6, Sep. 2018, Accessed: Jul. 29, 2021. [Online]. Available: <https://ejournal.stmik-wp.ac.id/index.php/icttech/article/view/30>.

- [2] Z. Zulfauzi and M. N. Alamsyah, "Penerapan Algoritma Naive Bayes Untuk Prediksi Penerimaan Mahasiswa Baru Studi Kasus Universitas Bina Insan Fakultas Komputer," *J. Teknol. Inf. Mura*, vol. 12, no. 02, pp. 156–165, Dec. 2020, doi: 10.32767/JTI.V12I02.1096.
- [3] R. K. Niswatin, "Sistem Seleksi Penerimaan Mahasiswa Baru Menggunakan Metode Weighted Product (WP)," in *Seminar Nasional Teknologi Informasi dan Multimedia 2016*, Feb. 2016, vol. 4, no. 1, pp. 1.3-31, Accessed: Aug. 20, 2021. [Online]. Available: <https://ojs.amikom.ac.id/index.php/semnasteknomedia/article/view/1338>.
- [4] I. Kurniawati, R. E. Indrajit, and M. Fauzi, "Peran Bussines Intelligence Dalam Menentukan Strategi Promosi Penerimaan Mahasiswa Baru," *IKRA-ITH Inform. J. Komput. dan Inform.*, vol. 1, no. 2, pp. 70–79, 2017, Accessed: Aug. 20, 2021. [Online]. Available: <https://journals.upi-yai.ac.id/index.php/ikraith-informatika/article/view/122>.
- [5] N. Yahya and A. Jananto, "Komparasi Kinerja Algoritma C.45 dan Naive Bayes untuk Prediksi Kegiatan Penerimaan Mahasiswa Baru (Studi Kasus : Universitas STIKUBANK Semarang)," in *Prosiding SENDI_U*, 2019, pp. 221–228, Accessed: Mar. 08, 2021. [Online]. Available: <https://www.unisbank.ac.id/ojs/index.php/sendu/article/view/7389>.
- [6] N. A. Haris, M. Abdullah, N. Hasim, and F. Abdul Rahman, "A study on students enrollment prediction using data mining," Jan. 2016, doi: 10.1145/2857546.2857592.
- [7] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques Third Edition*, Third Edit. Morgan Kaufmann, 2012.
- [8] B. Nakhkob and M. Khademi, "Predicted Increase Enrollment in Higher Education Using Neural Networks and Data Mining Techniques," *J. Adv. Comput. Res.*, vol. 7, no. 4, pp. 125–140, Nov. 2016, Accessed: Jul. 29, 2021. [Online]. Available: http://jacr.iausari.ac.ir/article_651013.html.
- [9] A. P. Dela Cruz *et al.*, "Higher education institution (Hei) enrollment forecasting using data mining technique," *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 9, no. 2, pp. 2060–2064, Mar. 2020, doi: 10.30534/IJATCSE/2020/179922020.
- [10] S. K. Wanjau and G. M. Muketha, "Improving Student Enrollment Prediction Using Ensemble Classifiers," *Int. J. Comput. Appl. Technol. Res.*, vol. 7, no. 3, pp. 122–128, 2018, Accessed: Jul. 29, 2021. [Online]. Available: <http://repository.mut.ac.ke/handle/123456789/2998>.
- [11] S. Shilbayeh and A. Abonamah, "Predicting Student Enrolments and Attrition Patterns in Higher Educational Institutions using Machine Learning," *Int. Arab J. Inf. Technol.*, vol. 18, no. 4, pp. 562–567, 2021, doi: 10.34028/18/4/8.
- [12] M. Mubassiran and M. I. Choldun, "Penerapan Algoritma K-Nearest Neighbors Untuk Menentukan Pola Penerimaan Mahasiswa Baru (Studi Kasus: Politeknik Pos Indonesia)," *Improve*, vol. 11, no. 2, pp. 1–8, Nov. 2019, Accessed: Jul. 29, 2021. [Online]. Available: <https://ejurnal.poltekpos.ac.id/index.php/improve/article/view/746>.
- [13] I. Loelianto, M. S. S. Thayf, and H. Angriani, "Implementasi Teori Naive Bayes Dalam Klasifikasi Calon Mahasiswa Baru STMIK Kharisma Makassar," *SINTECH (Science Inf. Technol. J.*, vol. 3, no. 2, pp. 110–117, Oct. 2020, doi: 10.31598/SINTECHJOURNAL.V3I2.651.
- [14] K. A. Pratama, G. A. Pradnyana, and I. K. R. Arthana, "Pengembangan Sistem Cerdas Untuk Prediksi Daftar Kembali Mahasiswa Baru Dengan Metode Naive Bayes (Studi Kasus: Universitas Pendidikan Ganesha)," *SINTECH (Science Inf. Technol. J.*, vol. 3, no. 1, pp. 22–34, Apr. 2020, doi: 10.31598/SINTECHJOURNAL.V3I1.523.
- [15] M. L. Dalafranka, "Penerapan Algoritma Id3 untuk Mendukung Sosialisasi Penerimaan Mahasiswa Baru yang Tepat Sasaran di UIN Raden Fatah Palembang," *Teknomatika*, vol. 10, no. 2, pp. 141–150, Oct. 2020, Accessed: Jul. 29, 2021. [Online]. Available: <http://ojs.palcomtech.com/index.php/teknomatika/article/view/501>.
- [16] A. Handayanto, K. Latifa, N. D. Saputro, and R. R. Waliansyah, "Analisis dan Penerapan Algoritma Support Vector Machine (SVM) dalam Data Mining untuk Menunjang Strategi Promosi," *JUITA J. Inform.*, vol. 7, no. 2, pp. 71–79, Nov. 2019, doi: 10.30595/JUITA.V7I2.4378.
- [17] S. Wulandari and M. Iqbal, "Analisis Kelayakan Lokasi Promosi Dalam Penerimaan Mahasiswa Baru (PMB) Dengan Algoritma Naive Bayes & Decission Tree C4.5," *KILAT*, vol. 10, no. 1, pp. 169–178, Apr. 2021, doi: 10.33322/KILAT.V10I1.1196.

- [18] “Weka 3 - Data Mining with Open Source Machine Learning Software in Java.” Waikato University, Accessed: Jul. 31, 2021. [Online]. Available: <https://www.cs.waikato.ac.nz/ml/weka/>.
- [19] D. T. Larose and C. D. Larose, *Data Mining and Predictive Analytics*. Hoboken, New Jersey: John Wiley & Sons, Inc., 2015.
- [20] B. Budiman, R. Nursyanti, R. Y. R. Alamsyah, and I. Akbar, “Data Mining Implementation Using Naïve Bayes Algorithm and Decision Tree J48 In Determining Concentration Selection,” *Int. J. Quant. Res. Model.*, vol. 1, no. 3, pp. 123–134, Sep. 2020, doi: 10.46336/ijqrm.v1i3.72.
- [21] M. M. Saritas and A. Yasar, “Performance Analysis of ANN and Naive Bayes Classification Algorithm for Data Classification,” *Int. J. Intell. Syst. Appl. Eng.*, vol. 7, no. 2, pp. 88–91, Jun. 2019, doi: 10.1039/b000000x.
- [22] “J48 numbers - Weka Wiki.” https://waikato.github.io/weka-wiki/not_so_faq/j48_numbers/ (accessed Mar. 08, 2021).
- [23] L. Breiman, “Random Forests,” *Mach. Learn. 2001 451*, vol. 45, no. 1, pp. 5–32, Oct. 2001, doi: 10.1023/A:1010933404324.
- [24] J. L. Speiser, M. E. Miller, J. Tooze, and E. Ip, “A comparison of random forest variable selection methods for classification prediction modeling,” *Expert Syst. Appl.*, vol. 134, pp. 93–101, Nov. 2019, doi: 10.1016/J.ESWA.2019.05.028.
- [25] F. Gorunescu, *Data Mining: Concepts, Models and Techniques*. Berlin: Springer-Verlag Berlin Heidelberg, 2011.