

PERBANDINGAN ALGORITMA KLASIFIKASI NAIVE BAYES DAN SVM PADA STUDI KASUS PEMBERIAN PENERIMA BEASISWA PPA

Safitri Linawati¹, Rizky Ade Safitri², Ahmad Rifqy Alfiyan³, Witriana Endah Pangesti⁴,
Monikka Nur Winarto⁵

^{1,3,5}STMIK Nusa Mandiri

e-mail: safitriw24@gmail.com, alfian7alfarisi@gmail.com, Monikkanw@gmail.com

^{2,4}Universitas Bina Sarana Informatika

e-mail: rizky.rzs@bsi.ac.id, witriana.weg@bsi.ac.id

Abstrak

Beasiswa PPA merupakan sebuah program beasiswa yang diberikan kepada mahasiswa aktif. Dalam melakukan seleksi penerima beasiswa PPA, akan ada tahapan-tahapan yang harus diproses untuk mendapatkan mahasiswa yang menerima beasiswa sesuai harapan. Data pendaftar beasiswa PPA dari tahun sebelumnya menjadi penunjang untuk keakuratan pengambilan keputusan dalam seleksi penerima beasiswa agar sesuai dengan kriteria yang telah ditetapkan. Untuk mendapatkan keakuratan pengambilan keputusan maka dibutuhkan *data mining* sebagai penunjang dalam pengambilan keputusan. Pada penelitian ini, penulis melakukan perbandingan 2 algoritma untuk mengetahui algoritma mana yang mempunyai keakuratan lebih tinggi dalam pengambilan keputusan penerima beasiswa PPA. Algoritma yang digunakan adalah algoritma klasifikasi *Naive Bayes* dan *Support Vector Machine* (SVM). Berdasarkan hasil komparasi antara algoritma *Naive Bayes* dan SVM (*Support Vector Machine*) yang dilakukan untuk mengklasifikasikan nilai akurasi tertinggi dengan 5 variabel dan jumlah data sebesar 122 dapat disimpulkan bahwa algoritma *Naive Bayes* memiliki tingkat akurasi lebih tinggi yaitu 90.90% dibandingkan dengan metode SVM yaitu 89.25%.

Kata kunci: *Beasiswa PPA, Klasifikasi, Naive Bayes, Support Vector Machine*

Abstract

PPA scholarship is a scholarship program given to active students. In selecting PPA scholarship recipients, there will be stages that must be processed to get students who receive scholarships as expected. PPA scholarship applicant data from the previous year is a support for the accuracy of decision making in the selection of scholarship recipients to fit the established criteria. To get the accuracy of decision making, data mining is needed to support decision making. In this study, the authors made a comparison of 2 algorithms to find out which algorithm has higher accuracy in decision making for PPA scholarship recipients. The algorithm used is the Naive Bayes classification algorithm and Support Vector Machine (SVM). Based on the results of the comparison between the Naive Bayes algorithm and SVM (Support Vector Machine) conducted to classify the highest accuracy value with 5 variables and the amount of data of 122 it can be concluded that the Naive Bayes algorithm has a higher accuracy level of 90.90% compared to the SVM method which is 89.25% .

Keywords: *Beasiswa PPA, Classification, Naive Bayes, Support Vector Machine*

1. Pendahuluan

Dalam Peraturan Pemerintah Nomor 48 tahun 2008 tentang Pendanaan Pendidikan, bagian 5, pasal 27 ayat (2), disebutkan bahwa Pemerintah dan Pemerintah Daerah sesuai dengan kewenangannya dapat memberi beasiswa kepada peserta didik yang berprestasi [Pedoman Beasiswa Peningkatan Prestasi Akademik (PPA). <http://belmawa.ristekdikti.go.id>. (diakses tanggal 22 Agustus 2017)]. Berdasarkan

Peraturan Pemerintah tersebut, maka Insitut Bisnis dan Multimedia asmi (IBM asmi) sebagai salah satu lembaga pendidikan tinggi melaksanakan program tersebut untuk memberikan beasiswa PPA kepada mahasiswa aktif.

Dalam melakukan seleksi penerima beasiswa PPA, akan ada tahapan-tahapan yang harus diproses untuk mendapatkan mahasiswa yang menerima beasiswa sesuai harapan. Data pendaftar beasiswa

PPA dari tahun sebelumnya menjadi penunjang untuk keakuratan pengambilan keputusan dalam seleksi penerima beasiswa agar sesuai dengan kriteria yang telah ditetapkan. Untuk mendapatkan keakuratan pengambilan keputusan maka dibutuhkan *data mining* sebagai penunjang dalam pengambilan keputusan.

Pada penelitian ini, penulis melakukan perbandingan 2 algoritma untuk mengetahui algoritma mana yang mempunyai keakuratan lebih tinggi dalam pengambilan keputusan penerima beasiswa PPA. Algoritma yang digunakan adalah algoritma klasifikasi *Naive Bayes* dan *Support Vector Machine* (SVM).

2. Metode Penelitian

2.1. Data mining

Data mining adalah proses yang menggunakan teknik statistic, matematika, kecerdasan buatan, dan machine learning untuk mengekstrasi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar (Hakim et al., 2017).

Istilah data mining memiliki hakikat sebagai disiplin ilmu yang tujuan utamanya adalah untuk menemukan, menggali, atau menambang pengetahuan dari data atau informasi yang kita miliki. Data mining, sering juga disebut sebagai Knowledge Discovery in Database (KDD). KDD adalah kegiatan yang meliputi pengumpulan, pemaknaan data, historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar (Mustafa et al., 2018).

2.2. Klasifikasi

Klasifikasi adalah salah satu pembelajaran yang paling umum di data mining. Klasifikasi didefinisikan sebagai bentuk analisis data untuk mengekstrak model yang akan digunakan untuk memprediksi label kelas. Kelas dalam klasifikasi merupakan atribut dalam satu set data yang paling unik yang merupakan variabel bebas dalam statistik. Klasifikasi data terdiri dari dua proses yaitu tahap pembelajaran dan tahap pengklasifikasian. Tahap pembelajaran merupakan tahapan dalam pembentukan model klasifikasi, sedangkan tahap pengklasifikasian merupakan tahapan penggunaan model klasifikasi untuk memprediksi label kelas dari suatu data (Sartika & Indra, 2017). Pada pengklasifikasian terdapat banyak algoritma yang dapat digunakan untuk

mengklasifikasikan data, namun dalam penelitian ini hanya akan membandingkan dua algoritma saja yaitu *Naive Bayes* dan *SVM (Support Vector Machine)*.

2.3. *naive Bayes*

Naive Bayes termasuk ke dalam pembelajaran *supervised*, sehingga pada tahapan pembelajaran dibutuhkan data awal berupa data pelatihan untuk dapat mengambil keputusan. Pada tahap pengklasifikasian akan dihitung nilai probabilitas dari masing-masing label kelas yang ada terhadap masukan yang diberikan. Label kelas yang memiliki nilai probabilitas paling besar yang akan dijadikan label kelas data masukan tersebut. *Naive Bayes* merupakan perhitungan *teorema bayes* yang paling sederhana, karena mampu mengurangi kompleksitas komputasi menjadi multiplikasi sederhana dari probabilitas [1].

$$P(C_i | X) = \frac{P(X | C_i)P(C_i)}{P(X)} \quad (1)$$

Keterangan :

X : Kriteria suatu kasus berdasarkan masukan

C_i : Kelas solusi pola ke-i, dimana i adalah jumlah label kelas

$P(C_i|X)$: Probabilitas kemunculan label kelas C_i dengan kriteria masukan X

$P(X|C_i)$: Probabilitas kriteria masukan X dengan label kelas C_i

$P(C_i)$: Probabilitas label kelas C_i

2.4. SVM (*Support Vector Machine*)

Support Vector Machine (SVM) adalah sistem pembelajaran yang pengklasifikasiannya menggunakan ruang hipotesis berupa fungsi-fungsi linear dalam sebuah ruang fitur (*feature space*) berdimensi tinggi, dilatih dengan algoritma pembelajaran yang didasarkan pada teori optimasi dengan mengimplementasikan *learning bias* yang berasal dari teori pembelajaran statistik (Widiastuti et al.,

2007). Support Vector Machine (SVM) termasuk dalam classifier yang dapat memprediksi probabilitas keanggotaan class dengan cara optimasi dan akan bekerja baik pada data dimensi tinggi.

2.5. WEKA

WEKA adalah sebuah paket tools machine learning praktis. WEKA merupakan singkatan dari Waikato Environment for Knowledge Analysis, yang dibuat di Universitas Waikato, New Zealand untuk penelitian, pendidikan dan berbagai aplikasi. WEKA mampu menyelesaikan masalah-masalah data mining di dunia nyata, khususnya klasifikasi yang mendasari pendekatan-pendekatan machine learning. Perangkat lunak ini ditulis dalam hirarki class Java dengan metode berorientasi objek dan dapat berjalan hampir di semua platform (Hakim et al., 2017).

3. Hasil dan Pembahasan

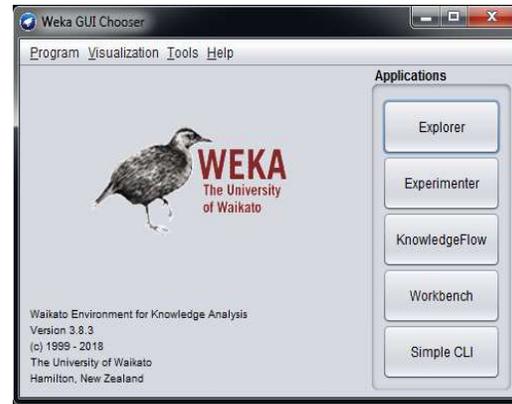
Penelitian yang dilakukan untuk membandingkan algoritma naïve bayes dan SVM adalah dengan mengumpulkan data pemberian beasiswa PPA. Sampel penelitian ini sebanyak 122 data. Data tersebut memiliki 5 variabel untuk menentukan kelas diterima atau ditolak. Variabel tersebut terdiri dari semester, pekerjaan, orang tua, penghasilan dan IPK. Setelah ekstraksi selesai kemudian diimport ke format CSV (Comma Separated Value, *CSV) kemudian diolah menggunakan tools weka. Data dalam format *csv dapat dianalisis menggunakan weka.

Dalam pengujian data ini menggunakan algoritma naïve bayes yang menggunakan tipe data kategorikal, kecuali SVM hanya dapat menggunakan tipe data numerik. Output yang dihasilkan mengandung persentase akurasi, presisi, dan recall serta terdapat confusion matrix yang menjadi pembandingan antara kedua algoritma tersebut.

3.1. Pengujian dengan WEKA

Pengujian dalam mengolah data dilakukan dengan menggunakan salah satu tools yaitu weka. Pada penelitian ini penulis menggunakan aplikasi weka 3.8. Adapun tahapan-tahapan dalam mengolah data menggunakan weka sebagai berikut:

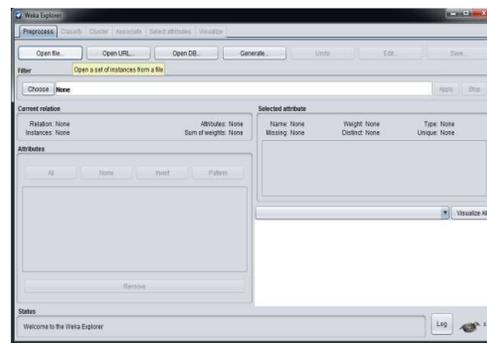
- a. Buka aplikasi weka kemudian pilih button explorer pada menu applications



Gambar 1. Tampilan Aplikasi Weka

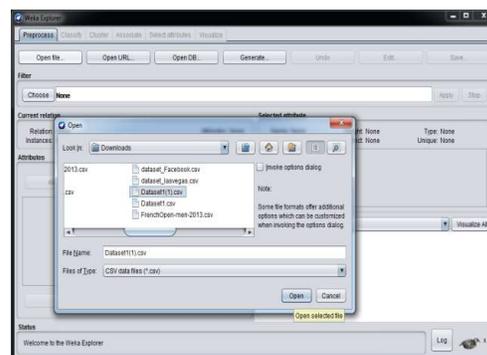
Gambar 1. Tampilan Aplikasi Weka

- b. pada menu preprocess pilih button open file untuk memasukkan dataset yang akan dihitung atau cari nilai akurasi.



Gambar 2. Tampilan untuk memasukkan Dataset

- c. Selanjutnya perhatikan tipe file dataset yang akan dihitung sesuaikan dengan data yang kita punya lalu klik button open.



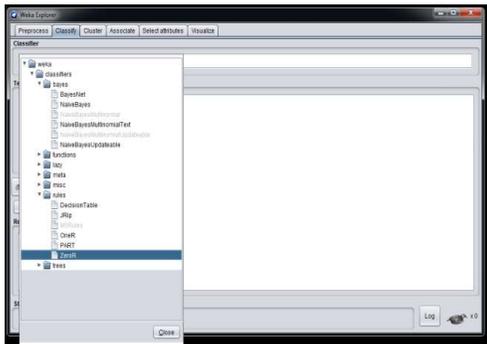
Gambar 3. Tampilan memilih dataset

d. ketika dataset berhasil dimasukkan maka akan tampil seperti ini.



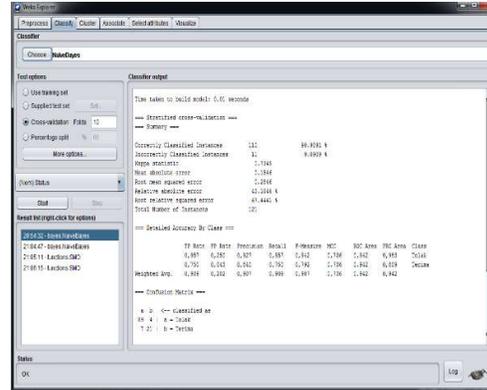
Gambar 4. Tampilan ketika dataset berhasil diinput

e. Tahapan Selanjutnya pada menu classify klik button chose untuk memilih algoritma apa yang akan digunakan dalam mengolah dataset. Klik folder bayes lalu klik algoritma naïve bayes kemudian close.



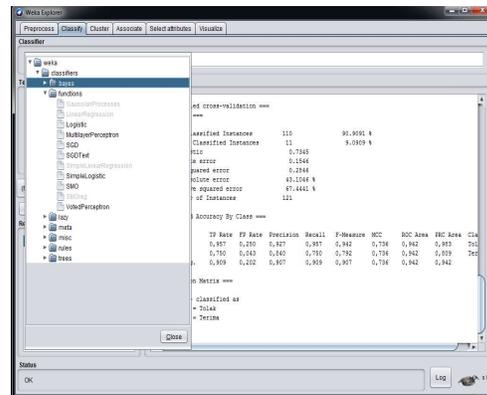
Gambar 5. Tampilan pemilihan algoritma Naive Bayes

f. tahapan selanjutnya klik tombol star maka akan otomatis berproses menghitung nilai akurasi serta confusion matrix nya.



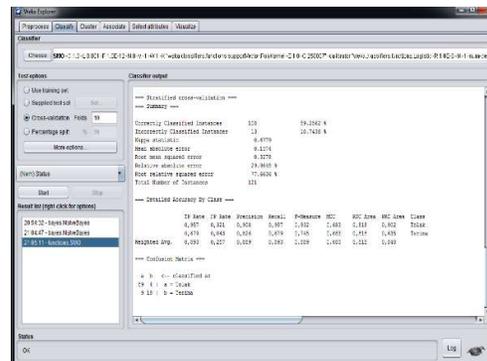
Gambar 6. Hasil mengolah data dengan algoritma Naive Bayes

g. Tahapan selanjutnya pada menu classify klik button choose untuk memilih algoritma apa yang akan digunakan dalam mengolah dataset. Klik folder function lalu klik algoritma SMO untuk SVM kemudian close.



Gambar 7. Tampilan mengolah data algoritma SVM

h. Tahapan selanjutnya klik tombol start maka akan otomatis berproses menghitung nilai akurasi serta confusion matrixnya.



Gambar 8. Hasil mengolah data dengan algoritma SVM

3.2. Hasil

Berdasarkan data yang telah dikumpulkan yaitu 122 sampel yang dijadikan dataset yang diolah menggunakan microsoft excel dikonversi menjadi format csv agar dapat diproses melalui aplikasi weka. Dataset dengan format file csv kemudian diproses menggunakan 2 algoritma untuk melakukan perbandingan akurasi yang lebih tinggi dalam menentukan penerima beasiswa PPA yaitu algoritma Naïve bayes dan SVM pada aplikasi weka. Maka dapatlah hasilnya sebagai berikut:

Tabel 1. Hasil Pengolahan data dengan algoritma naïve bayes

Naïve bayes	
	Percentage Split
Correctly Calssified	90,9091 %
Incorrectly Clasified	9,0909 %
Kappa Statistic	0,7345
Mean Absolute Error	0,1546
Root Mean Squared Error	0,2846
Relative Absolute Errorr	43,1046 %
Root Relative Squared Error	67,4441%

Tabel 2 Hasil Pengolahan data dengan algoritma SVM

Naïve bayes	
	Percentage Split
Correctly Calssified	89,2562 %
Incorrectly Clasified	10,7438 %
Kappa Statistic	0,6779
Mean Absolute Error	0,1074
Root Mean Squared Error	0,3278
Relative Absolute Errorr	29,9648 %
Root Relative Squared Error	77,6636 %

Dari tabel diatas dapat disimpulkan bahwa algoritma naïve bayes memiliki tingkat persentasi tertinggi yaitu sebesar 90,9091 %

4. Kesimpulan

Perbandingan algoritma dengan menggunakan WEKA 3.8 dapat dilihat dari beberapa nilai yang dihasilkan antara lain *correctly classified instance*, *incorrectly* Tabel 2. Hasil Pengolahan data dengan algoritma SVM (*Support Vector Machine*)

classified, *kappa statistic*, *mean absolute error*, *root mean squared error*, *relative absolute error*, *root relative squared error*. Selain nilai-nilai tersebut, *confusion matrix* yang dihasilkan dapat diketahui jumlah *false positive* dan *false negative* sehingga dapat dihitung performa masing-masing algoritma untuk masingmasing kelas.

Berdasarkan hasil komparasi antara algoritma *Naïve Bayes* dan SVM (*Support Vector Machine*) yang dilakukan untuk mengklasifikasikan nilai akurasi tertinggi dengan 5 variabel dan jumlah data sebesar 122 dapat disimpulkan bahwa algoritma *Naïve Bayes* memiliki tingkat akurasi lebih tinggi yaitu 90.90% dibandingkan dengan metode SVM yaitu 89.25%. Meskipun demikian dilihat dari nilai presisi dan recall untuk masing-masing metode hanya memiliki selisih yang tidak jauh berbeda. Hal ini menunjukkan bahwa untuk jumlah fitur/variabel yang sama akan menghasilkan nilai akurasi yang tidak jauh berbeda.

Referensi

- Hakim, O. L., Ag, S., & Si, M. (2017). *Kabupaten Karawang*. 2(2), 43–53. <https://journal.unsika.ac.id/index.php/politikomindonesiana/article/view/963>
- Mustafa, M. S., Ramadhan, M. R., & Thenata, A. P. (2018). Implementasi Data Mining untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier. *Creative Information Technology Journal*, 4(2), 151. <https://doi.org/10.24076/citec.2017v4i2.106>
- Sartika, D., & Indra, D. (2017). Perbandingan Algoritma Klasifikasi Naive Bayes, Nearest Neighbour, dan Decision Tree pada Studi Kasus Pengambilan Keputusan Pemilihan Pola Pakaian. *Jurnal Teknik Informatika Dan Sistem Informasi*, 1(2), 151–161.
- Widiastuti, D., Informasi, J. S., & Gunadarma, U. (2007). *Analisa Perbandingan Algoritma Svm , Naive Bayes , Dan Decision Tree Dalam Mengklasifikasikan Serangan (Attacks)*. 1–8.