JURNAL PINTER
VOL 2. NO.2 DESEMBER 2018
e-ISSN: 2597- 4475
http://doi.org/10.21009/pinter.2.2.4

Perbandingan Algoritma Naive Bayes Dan Support Vector Machine Dalam Seleksi Kelulusan Pemberkasan Beasiswa BPP-PPA Fakultas Teknik Universitas Negeri Jakarta

Fakhriyani[1], Widodo[2], Bambang Prasetya Adhi[3]

[1,2,3] Pendidikan Teknik Informatika dan Komputer Fakultas Teknik
Universitas Negeri Jakarta

[1]fanybroto@gmail.com, [2]widodo@unj.ac.id, [3]bambangpadhi@unj.ac.id

ABSTRAK

Beasiswa merupakan salah satu program untuk membantu meringankan mahasiswa dalam membayar uang kuliah, namun sering terjadi kesalahan dalam pemberian beasiswa tersebut karena masih dilakukan secara manual dan tidak adanya kriteria yang jelas bagaimana seorang mahasiswa dapat memperoleh beasiswa. Untuk mengantisipasi agar tidak terjadinya kesalahan dalam pemberian beasiswa maka dibutuhkan sebuah Sistem Pendukung Keputusan, namun sebelum dilakukan pembuatan sistem tersebut dirasa perlu untuk mengetahui algoritma terbaik untuk menyeleksi berkas beasiswa tersebut. Penelitian ini menggunakan dua algoritma *Data Mining* yaitu algoritma *Naïve Bayes* dan *Support Vector Machine*. *Naïve Bayes* merupakan metode pengklasifikasian yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu *class* berdasarkan pengalaman di masa sebelumnya dengan kondisi antar atribut saling bebas. *Support Vector Machine* adalah sebuah metode prediksi dalam klasifikasi yang dapat dilakukan pada kasus yang secara linier dapat dipisahkan, maupun non-linier dengan menggunakan konsep kernel pada ruang kerja berdimensi tinggi. Data mahasiswa yang lulus dan tidak lulus seleksi berkas beasiswa BPP-PPA akan diolah menggunakan algoritma *Naïve Bayes* dan *Support Vector Machine*. Setelah diklasifikasi kedua algoritma tersebut akan dihitung hasil akurasinya menggunakan *K-fold Cross Validation*. Berdasarkan hasil contoh kasus seleksi menunjukan bahwa hasil perhitungan akurasi algoritma *Naïve Bayes* adalah 0.7542, sedangkan hasil akurasi algoritma *Support Vector Machine* adalah 0.99. Kedua sistem telah mampu menangani proses penyeleksian kelulusan pemberkasan beasiswa BPP-PPA Fakultas Teknik Universitas Negeri Jakarta. Algoritma *Support Vector Machine* menghasilkan rata-rata akurasi 0.99 yang mendekati 1, maka algoritma tersebut dinilai lebih akurat dan direkomendasikan untuk penelitian selanjutnya.

Kata Kunci: Data Mining, Beasiswa, Akurasi, Naïve Bayes, Support Vector Machine.

1. Pendahuluan

Latar Belakang

Beasiswa adalah tunjangan uang yang diberikan kepada pelajar atau mahasiswa sebagai bantuan biaya belajar[1]. Pemberian beasiswa dapat dikategorikan sebagai pemberian cuma-cuma maupun pemberian pada ikatan kerja (ikatan dinas) setelah selesainya pendidikan. Tujuan dari diberikannya beasiswa adalah untuk membantu para pelajar atau mahasiswa dalam mengantisipasi mahalnya memperoleh pendidikan sehingga pelajar tersebut dapat memenuhi segala kebutuhan dalam proses belajar agar pendidikan dapat terlaksana dengan baik. Beasiswa dapat diberikan oleh lembaga pemerintah, perusahaan, ataupun yayasan.

Salah satu penyelenggara beasiswa di lembaga pemerintah adalah Direktorat Jenderal Pembelajaran dan Kemahasiswaan Kementerian Riset, Teknologi, dan Pendidikan Tinggi. Program beasiswa yang

[1] Departemen Pendidikan Nasional, K*amus Besar Bahasa Indonesia Edisi Ketiga*. (Jakarta: Balai Pustaka,2005), hlm. 89.

Avalaiable at:
http://journal.unj.ac.id/unj/index.php/pinter/article/view/16217

dibeikan oleh DIKTI adalah Beasiswa Peningkatan Prestasi Akademik (Beasiswa-PPA) dan Beasiswa Bantuan Biaya Pendidikan Peningkatan Prestasi Akademik (BPP-PPA). Bantuan biaya pendidikan tersebut diberikan kepada mahasiswa yang orang tuanya tidak mampu untuk membiayai pendidikannya, dan kepada mahasiswa yang mempunyai prestasi tinggi, baik kurikuler maupun ekstrakulikuler[2].

Universitas Negeri Jakarta menjadi salah satu universitas yang bekerja sama dengan DIKTI dalam menyelenggarakan beasiswa BPP-PPA. Saat ini fakultas teknik masih menggunakan cara manual untuk menentukan mahasiswa yang berhak menerima beasiswa BPP-PPA. Sehingga rawan akan kesalahan seperti *human error* dan pemberian beasiswa tidak tepat sasaran. Kemudian pengolahan data kurang efektif, membutuhkan waktu yang relatif lama, dan sering terjadi subjektifitas dari para pengambil keputusan ataupun kecurangan-kecurangan lain yang dapat terjadi pada pemberian beasiswa BPP-PPA.

Agar tidak terjadi kesalahan dan untuk mempermudah para petugas fakultas dalam menyeleksi dan menentukan mahasiswa yang berhak menerima beasiswa BPP-PPA, maka diperlukan sebuah Sistem Pendukung Keputusan yang berfungsi untuk membantu melakukan seleksi kepada para calon penerima beasiswa. Namun sebelum dibuatnya Sistem Pendukung Keputusan diperlukan perencanaan yang baik, salah satunya dengan pemilihan algoritma yang digunakan oleh sistem nanti.

Terdapat banyak algoritma pada *Data Mining* yang bisa digunakan pada penelitian ini, namun diantara banyaknya algoritma tersebut penulis memilih dua algoritma yaitu algoritma *Naïve Bayes* dan algoritma *Support Vector Machine*. Penulis memilih untuk membandingkan algoritma *Naïve Bayes* dan *Support Vector Machine* dikarenakan kedua algoritma tersebut termasuk dalam dua dari sepuluh algoritma terbaik.

Alasan penulis membandingkan kedua algoritma tersebut karena belum ada penelitian yang menggunakan algoritma *Naïve Bayes dan Support Vector Machine* untuk kasus seleksi berkas beasiswa BPP-PPA. Kedua algoritma tersebut penulis pilih karena merupakan algoritma klasifikasi pada *Data Mining*. Kemudian kedua algoritma tersebut menggunakan dua teknik yang berbeda, untuk algoritma *Naïve Bayes* menggunakan teknik statistik sedangkan *Support Vector Machine* menggunakan teknik *machine learning*.

Perbandingan algoritma *Naïve Bayes* dan *Support Vector Machine* dilakukan untuk mencari algoritma yang lebih akurat pada kasus seleksi berkas beasiswa BPP-PPA. Penulis menggunakan data pendaftar dan penerima beasiswa BPP-PPA Fakultas Teknik Universitas Negeri Jakarta tahun 2011 hingga tahun 2015 untuk kepentingan perbandingan algoritma. Setelah diketahui algoritma yang lebih akurat dan tepat untuk seleksi berkas beasiswa BPP-PPA diharapkan dapat membantu pembuatan Sistem Pendukung Keputusan pada penelitian selanjutnya.

Berdasarkan penjelasan diatas maka Penulis tertarik untuk mencoba memecahkan masalah yang ada khususnya di bidang penyeleksian berkas beasiswa. Hal ini yang mendorong Penulis untuk mencoba menganalisa "Perbandingan Algoritma *Naïve Bayes* dan *Support Vector Machine* dalam Seleksi Kelulusan Pemberkasaan Beasiswa BPP-PPA Fakultas Teknik Universitas Negeri Jakarta"

2. Dasar Teori

2.1. *Data Mining*

Data Mining adalah proses menemukan pola yang menarik dan pengetahuan dari sejumlah data yang besar. Sumber data bisa termasuk databases, data *warehouses*, *theWeb*, repositori informasi lainnya, atau data yang mengalir ke dalam sistem secara dinamis[3].

Data Mining merupakan teknologi yang menggabungkan metoda analisis tradisional dengan algoritma yang canggih untuk memproses data dengan volume besar. *Data mining* adalah suatu istilah yang digunakan untuk menemukan pengetahuan yang tersembunyi di dalam *database*. Data mining merupakan proses semi otomatik yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi pengetahuan potensial dan berguna yang bermanfaat yang tersimpan di dalam database besar[4].

Berdasarkan dari teori para ahli diatas dapat disimpulkan bahwa *Data Mining* adalah sebuah teknik untuk mencari pola dalam sebuah kumpulan data yang sangat besar. Teknik yang digunakan antara lain teknik statistik, matematika, kecerdasan buatan dan *machine learning*. *Data Mining* menganalisis kumpulan data yang sebelumnya hanya disimpan pada database menjadi sebuah pengetahuan yang bermanfaat.

2.2. Klasifikasi

Klasifikasi adalah suatu bentuk analisis data yang mengekstrak model dan menggambarkan kelas

[2] Direktorat Jendral Pendidikan Tinggi, *Pedoman Umum Beasiswa Bantuan Biaya Pendidikan Peningkatan Prestasi Akademik (PPA), (*Direktorat Jendral Pembelajaran dan Kemahasiswaan Kementerian Riset, Teknologi, dan Pendidikan Tinggi, 2015), hlm. i.

[3] Jiawei Han, dkk., *Data Mining Concepts and Techniques, 3rd ed*. (Waltham, USA: Elsevier Inc., 2012), hlm. 8.

[4] Turban E, dkk., *Decicion Support Systems and Intelligent Systems.* (Yogyakarta:Andi Offset,2005), hlm. 23.

Avalaiable at:
http://journal.unj.ac.id/unj/index.php/pinter/article/view/16217

data penting. Seperti model-model, disebut *classifiers*, memprediksi kategoris (tidak beraturan) kelas label[5].

Klasifikasi data adalah dua tahap proses, terdiri dari tahap belajar *(learning step*) dimana model klasifikasi dibangun dan tahap klasifikasi dimana model yang digunakan untuk memprediksi label kelas untuk data yang diberikan[6].

Berdasarkan teori para ahli diatas dapat disimpulkan klasifikasi adalah proses pengelompokkan suatu data ke dalam beberapa kelas data. Pengelompokkan data tersebut dilakukan melalui tahap pembelajaran dimana data-data tersebut dianalisis, kemudian tahap klasifikasi untuk memisahkan data ke dalam kelas yang dianggap sesuai dengan kriteria data tersebut.

Pada dasarnya klasifikasi dibagi menjadi dua, *rule based classification* dan *non rule based classification*. Pada penulisan kali ini salah satu algoritma yang digunakan adalah *Naïve Bayes* yang termasuk dalam salah satu *non rule based classification.*

Naïve Bayes merupakan salah satu penerapan teorama *Bayes*. *Naïve Bayes* didasarkan pada asumsi penyederhanaan bahwa nilai atribut secara kondisional saling bebas jika diberikan nilai output[7].

Naïve Bayes adalah pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu *class*[8].

Formula umum untuk prediksi *Bayes* didasari oleh teorama *Bayes* adalah sebagai berikut :

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)}$$

Keterangan :

X : Data dengan *class* yang belum diketahui
H : Hipotesis data X merupakan suatu class spesifik
P(H|X) : Probabilitas hipotesis H berdasar kondisi X (*posteriori probability*)
P(H) : Probabilitas hipotesis H (*prior probability*)
P(X|H) : Probabilitas X berdasarkan kondisi pada hipotesis H
P(X) : Probabilitas X

[5] Jiawei Han, dkk., *Data Mining Concepts and Techniques, 3rd ed*. (Waltham, USA: Elsevier Inc., 2012), hlm. 327.

[6] Ibid, h. 328.

[7] Khafiizh Hastuti, *Analisis Komparasi Algoritma Klasifikasi Data Mining untuk Prediksi Mahasiswa Non Aktif*. (Seminar Nasional Teknologi Informasi dan Komunikasi Terapan, 2012), hlm. 243.

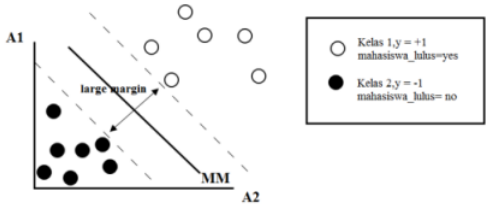
[8] Eko Prasetyo, *Data Mining : Konsep dan Aplikasi menggunakan MATLAB, 1st ed*. (Yogyakarta:Andi Offset, 2012), hlm. 37.

2.3. *Support Vector Machine*

SVM adalah sebuah metode untuk klasifikasi baik data *linear* maupun *nonlinear*. SVM merupakan algoritma yang bekerja sebagai berikut. Menggunakan pemetaan *nonlinear* untuk mengubah data pelatihan asli ke dimensi yang lebih tinggi. Dalam dimensi baru ini, akan mencari *linear* yang optimal untuk memisahkan *hyperplane* (yaitu, "batas keputusan" yang memisahkan tupel dari satu kelas ke kelas lainnya). Dengan pemetaan *nonlinear* yang tepat untuk dimensi yang cukup tinggi, data dari dua kelas selalu dapat dipisahkan dengan *hyperplane*. SVM menentukan *hyperplane* menggunakan vektor dukungan (tupel pelatihan "penting") dan *margin* (didefinisikan oleh vektor dukungan)[9].

SVM memiliki prinsip dasar *linear classifier* yaitu kasus klasifikasi yang secara linier dapat dipisahkan, namun SVM telah dikembangkan agar dapat bekerja pada problem *nonlinear* dengan memasukkan konsep kernel pada ruang kerja berdimensi tinggi. Pada ruang berdimensi tinggi, akan dicari *hyperplane* yang dapat memaksimalkan jarak (*margin*) antara kelas data[10].

Berdasarkan teori parah ahli diatas dapat disimpulkan *Support Vector Machine* adalah sebuah metode klasifikasi untuk data *linear* maupun *nonlinear* menggunakan *hyperlane* pembatas. Pada kasus data *nonlinear* konsep yang digunakan adalah konsep kernel dengan ruang kerja berdimensi tinggi. *Hyperlane* pembatas terbaik dapat ditentukan dengan mencari *support vector* dan *maximum margin*, kemudian data tersebut akan dipisahkan sesuai dengan *hyperlane* yang sudah ditentukan.



Gambar 2.1 Pemisahan 2 Kelas Data dengan *Maximum Margin*

Pada Gambar 2.1, SVM menemukan *hyperplane* pemisah maksimum, yaitu *hyperplane* yang mempunyai jarak maksimum antara tupel pelatihan terdekat. *Support vector* ditunjukkan dengan batasan tebal pada titik tupel.

[9] Jiawei Han, dkk., *Data Mining Concepts and Techniques, 3rd ed*. (Waltham, USA: Elsevier Inc., 2012), hlm. 408.

[10] Pusphita Anna, *Penerapan Metode Klasifikasi Support Vector Machine (SVM) pada Data Akreditasi Sekolah Dasar (SD) di Kabupatern Magelang*. (Seminar Nasional Informatika, 2014), hlm. 812.

Avalaiable at:
http://journal.unj.ac.id/unj/index.php/pinter/article/view/16217

2.4. Beasiswa Bantuan Biaya Pendidikan Peningkatan Prestasi Akademik (BPP-PPA)

Sejak tahun 2012 istilah Beasiswa Peningkatan Prestasi Akademik (PPA) dan Bantuan Belajar Mahasiswa (BBM) disesuaikan dengan istilah yang sejalan dengan ketentuan yang ada yaitu menjadi Beasiswa Peningkatan Prestasi Akademik (Beasiswa-PPA) dan Bantuan Biaya Pendidikan Peningkatan Prestasi Akademik (BPP-PPA). Beasiswa Bantuan Biaya Pendidikan Peningkatan Prestasi Akademik (BPP-PPA) adalah salah satu beasiswa yang diberikan oleh Pemerintah melalui Direktorat Jenderal Pembelajaran dan Kemahasiswaan Kementerian Riset, Teknologi, dan Pendidikan Tinggi. Bantuan biaya pendidikan tersebut diberikan kepada mahasiswa yang orang tuanya tidak mampu untuk membiayai pendidikannya, dan kepada mahasiswa yang mempunyai prestasi tinggi, baik kurikuler maupun ekstrakulikuler[11].

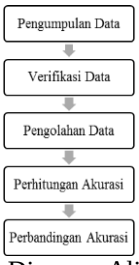
2.5. *K-fold Cross Validation*

Bentuk umum pendekatan ini disebut dengan *K-fold Cross Validation*, yang memecah set data menjadi k bagian set data dengan ukuran yang sama. Setiap kali berjalan, satu pecahan berperan sebagai set data uji sedangkan pecahan lainnya menjadi set data latih. Prosedur tersebut dilakukan sebanyak k kali sehingga setiap data berkesempatan menjadi data uji tepat satu kali dan menjadi data latih sebanyak k-1 kali. Total *error* didapatkan dengan menjumlah semua *error* yang didapatkan dari k kali proses[12].

Dataset training secara acak dibagi menjadi K *dataset* dengan ukuran yang sama dimana setiap bagian memiliki kurang lebih distribusi *class* yang sama. *Classifier* dilatih k-kali, setiap waktu dengan set yang berbeda dan dikeluarkan untuk dataset tes. Diperkirakan kesalahan adalah *mean* dari k-kesalahan[13].

Berdasarkan teori para ahli diatas dapat disimpulkan bahwa dalam teknik *K-fold Cross Validation dataset* dibagi menjadi 10 partisi secara acak, kemudian dilakukan sejumlah 10 kali eksperimen dimana masing-masing eksperimen menggunakan data partisi ke-10 sebagai *data teseting* dan memanfaatkan sisa partisi lainnya sebagai *data training*. Untuk 1 iterasi dilakukan 10 kali pengulangan (1-st fold, 2-nd fold,……10-th fold) dengan 10 klasifikasi dimana 1 *data testing* dan 9 *data training* secara bergantian.

3. Metodologi

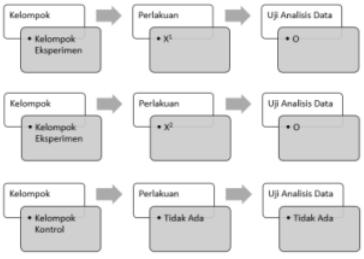


Gambar 3.1 Diagram Alir Penelitian

Berdasarkan Gambar 3.1 Langkah pertama dalam penelitian adalah pengumpulan data. Penulis mendapatkan *soft copy* 500 data mahasiswa yang lulus dan 200 data mahasiswa tidak lulus seleksi pemberkasan beasiswa BPP-PPA Fakultas Teknik Universitas Negeri Jakarta tahun 2011-2015.

Setelah data terkumpul dan mencukupi kebutuhan penelitian maka langkah selanjutnya adalah verifikasi data. Pada tahap verifikasi data ini terdapat tiga proses yaitu, pertama proses perubahan data sesuai parameter, kedua proses perubahan data ke dalam bentuk vektor, proses ketiga adalah pemberian label pada masing-masing vektor data. Untuk data mahasiswa yang lulus seleksi berkas diberikan label angka 1, sedangkan untuk data mahasiswa yang tidak lulus seleksi berkas diberikan label angka 0. Setelah 3 proses tersebut dilakukan maka akan masuk ke tahap selanjutnya yaitu tahap pengolahan data.

Pada tahap pengolahan data penulis merancang penelitian yang akan dilakukan seperti ilustrasi pada Gambar 3.2 berikut.



Gambar 3.2 Rancangan Penelitian

Keterangan :
KE : Kelompok Eksperimen
KK : Kelompok Kontrol
X^1 : Klasifikasi data menggunakan algoritma *Naïve Bayes*
X^2 : Klasifikasi data menggunakan Algoritma SVM
O : Evaluasi hasil akurasi algoritma menggunakan metode *K-fold Cross Validation*

[11] Direktorat Jendral Pendidikan Tinggi, *Pedoman Umum Beasiswa Bantuan Biaya Pendidikan Peningkatan Prestasi Akademik (PPA), (*Direktorat Jendral Pembelajaran dan Kemahasiswaan Kementerian Riset, Teknologi, dan Pendidikan Tinggi, 2015), hlm. i.

[12] Eko Prasetyo, *DATA MINING – Mengolah Data menjadi Informasi Menggunakan Matlab*, (Yogyakarta: Andi Offset, 2014), hlm. 264.

[13] Ron Kohavi, *A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. (International Joint Conference on Artifical Intellegence*), hlm. 2.

Avalaiable at:
http://journal.unj.ac.id/unj/index.php/pinter/article/view/16217

Berdasarkan Gambar 3.2 diatas, penelitian ini melibatkan dua kelompok data yaitu kelompok eksperimen dan kelompok kontrol. Hanya satu kelompok saja yang diberikan perlakuan yaitu kelompok eksperimen yang diolah menggunakan dua algoritma (algoritma *Naïve Bayes* dan *Support Vector Machine*). Sedangkan kelas kontrol tidak diolah menggunakan dua algoritma atau tidak diberikan perlakuan apapun.

Baik kelas eksperimen maupun kelas kontrol keduanya berasal dari sumber data yang sama, yaitu data mahasiswa yang lulus maupun tidak lulus seleksi berkas beasiswa BPP-PPA tahun 2011-2015 di Fakultas Teknik Universitas Negeri Jakarta. Perbedaannya adalah pada kelas eksperimen data yang sama akan diproses menggunakan dua algoritma yang berbeda.

4. Hasil dan Analisis

4.1 Deskripsi Hasil Penelitian

Data pendaftar beasiswa BPP-PPA yang dikumpulkan oleh penulis sebanyak 500 data mahasiswa yang lulus pemberkasan dan 200 data mahasiswa yang tidak lulus pemberkasan. Adapun syarat dan kriteria yang penulis simpulkan melalui data yang didapatkan, buku pedoman beasiswa BPP-PPA serta kebijakan Universitas Negeri Jakarta untuk menentukan lulus atau tidaknya mahasiswa pendaftar beasiswa BPP-PPA adalah sebagai berikut:

a. IPK minimal 2,75
b. Gaji orang tua minim atau kecil
c. Rekening listrik dibawah 2200v
d. Surat Keterangan Tidak Mampu (SKTM) dari pihak kelurahaan yang akan mempengaruhi kelengkapan berkas dari mahasiswa tersebut.

4.1.1. Verifikasi Data

Sebelum data diproses menggunakan algoritma *Naïve Bayes* dan *Support Vector Machine* maka data tersebut harus diverifikasi terlebih dahulu. Verifikasi data dilakukan guna mengurangi kesalahan pemrosesan data pada saat dilaksanakannya eksperimen dan format data sesuai dengan kebutuhan penelitian.

4.1.1.1. Proses Perubahan Data Sesuai Parameter

Tabel 4.1 Potongan Data Inti yang Sudah Dipersempit

No	N a m a	IPK	Daya Listrik (Volt)	Penerbit SKTM	Gaji
1	Fathudin	2.79	1300	Lurah	Rp 1,500,000
2	Okky Andri Yanto	3.25	1300	Lurah	Rp 1,500,000
3	Farizal	2.95	450	Lurah	Rp 1,000,000
4	Mohamad Azhar	2.84	1300	Lurah	Rp 2,400,000
5	Bagus Sopyan	3.11	900	Lurah	Rp 2,900,000
6	Ika Yunsita Pratiwi	3,00	1300	Lurah	Rp 1,500,000
7	Yulianti Suryani	3.35	900	Lurah	Rp 1,000,000
8	Yudi Hastomi	3.24	450	Lurah	Rp 1,800,000
9	Selly Yunis	3.13	450	Lurah	Rp 800,000
10	Teguh Imam Perdana	3.06	900	Lurah	Rp 800,000

Tabel 4.1 adalah data utama yang diperlukan dalam proses eksperimen yaitu data nama, IPK, daya listrik, penerbit SKTM (salah satu indikator kelengkapan berkas) dan gaji orang tua. Data utama tersebut yang nantinya akan dijadikan parameter oleh algoritma *Naïve Bayes* dan *Support Vector Machine* pada saat pengolahan data.

4.1.1.2. Perubahan Data ke dalam Bentuk Vektor

Data yang sudah sesuai dengan parameter yang dibutuhkan untuk penelitian akan diubah ke dalam bentuk vektor guna mempermudah proses kerja algoritma.

Tabel 4.2 Inisialisasi Data Menjadi Nominal Kecil

IPK	:	0 – 2,74	→	1
	:	2,75 – 3,25	→	2
	:	3,26 – 4,00	→	3
Gaji Orang Tua	:	Rp. 0 – Rp. 2.000.000	→	1
	:	Rp. 2.000.001 – Rp. 5.000.000	→	2
	:	Rp. 5.000.001 – Rp. 10.000.000	→	3
Berkas	:	Lengkap	→	1
	:	Tidak Lengkap	→	2
Rekening Listrik	:	450 V	→	1
	:	900 V	→	2
	:	1300 V	→	3
	:	2200 V	→	4

Tabel 4.2 adalah inisialisasi yang penulis tentukan untuk merubah data mahasiswa kedalam bentuk nominal yang lebih kecil.

```
X =
[[2,1,1,3],[2,1,1,1],[2,1,1,3],[2,1,1,3],[2,1,1,2],[2,1,1,1],[3,2,1,3],[3,1,1,2],[3,1,
1,1],[2,2,1,2],

[2,1,1,3],[2,2,1,3],[2,1,1,2],[2,1,1,3],[3,1,1,2],[2,1,1,3],[2,1,1,2],[2,1,1,2],[2,1,1
,3],[2,1,1,2],

[2,1,1,3],[2,2,1,1],[3,1,1,2],[2,1,1,1],[2,2,1,1],[2,2,1,2],[2,1,1,1],[2,2,1,3],[2,1,1
,1],[2,2,1,2],

[2,2,1,3],[2,2,1,3],[2,1,1,3],[2,2,1,2],[2,1,1,2],[3,2,1,3],[2,2,1,3],[2,1,1,2],[2,2,1
,1],[2,1,1,1],

[3,1,1,3],[3,2,1,3],[2,1,1,2],[2,1,1,1],[2,1,1,2],[2,1,1,3],[3,1,1,3],[3,1,1,3],[3,2,1
,3],[2,2,1,3],

[3,2,1,3],[2,2,1,3],[2,2,1,3],[2,2,1,3],[2,2,1,3],[2,1,1,3],[2,1,1,2],[2,1,1,3],[2,1,1
,2],[2,1,1,3],

[2,1,1,2],[2,1,1,3],[2,1,1,1],[2,1,1,1],[2,1,1,3],[2,1,1,1],[3,2,1,3],[2,1,1,2],[3,2,1
,1],[2,2,1,1],

[2,2,1,2],[2,1,1,1],[3,1,1,3],[2,1,1,3],[2,2,1,3],[2,1,1,2],[2,2,1,3],[2,1,1,2],[2,1,1
,3],[2,2,1,3],

[3,1,1,1],[3,2,1,2],[3,1,1,1],[3,1,1,2],[3,2,1,1],[3,2,1,3],[2,2,1,2],[2,2,1,1],[2,1,1
```

Gambar 4.1 Data yang Telah Diubah Menjadi Bentuk Vektor

Gambar 4.1 adalah data yang sudah diubah ke dalam bentuk vektor dengan urutan [ipk, gaji orang tua, kelengkapan berkas, rekening listrik]. Data mahasiswa yang lulus maupun tidak lulus sebanyak 700 data dan telah dibagi menjadi 10 kelompok, dimana dalam 1 kelompok terdapat 70 data mahasiswa dengan urutan 50 data mahasiswa yang lulus dan 20 data mahasiswa tidak lulus. Pengelompokkan data ini berguna untuk tahap evaluasi akurasi algoritma menggunakan K-fold Cross Validation.

4.1.1.3. Pemberian Label Pada Data

Setelah data diubah kedalam bentuk vektor maka tahap selanjutnya untuk memudahkan proses

Avalaiable at:
http://journal.unj.ac.id/unj/index.php/pinter/article/view/16217

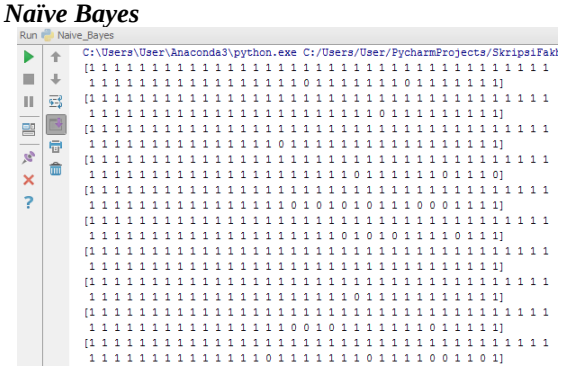
pembelajaran data (*learning step*) maka data tersebut harus diberikan label. Untuk data mahasiswa yang lulus pemberkasan beasiswa akan dilabeli dengan angka 1, sedangkan untuk data mahasiswa yang tidak lulus pemberkasan beasiswa akan dilabeli dengan angka 0.

Label data tersebut kemudian dibagi menjadi dua kelompok yaitu kelompok label *data train* untuk membantu algoritma mempelajari data, dan kelompok label *data test* untuk mendapatkan hasil akurasi yaitu membandingkan hasil klasifikasi oleh algoritma dengan hasil yang seharusnya. Untuk satu kelompok data dengan jumlah 70 data diberi urutan label 50 data lulus (1) dan 20 data tidak lulus (0). Karena dalam 1 iterasi terdapat 9 kelompok data train maka terdapat 630 label data train, dan 1 kelompok data test sebanyak 70 label data test.

4.2. Analisis Data Penelitian

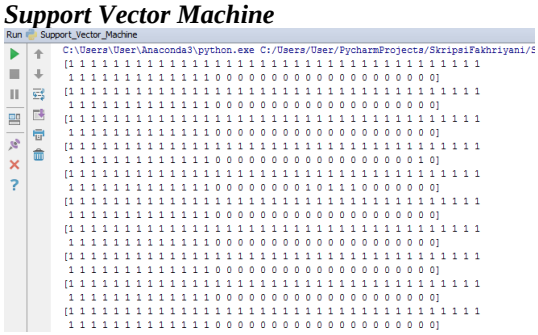
Data telah di verifikasi dan selanjutnya akan masuk ke dalam tahap pengolahan data. Kumpulan data dalam bentuk vektor tersebut akan diolah menggunakan dua algoritma yaitu *Naïve Bayes* dan *Support Vector Machine*. Instrumen atau alat yang digunakan oleh peneliti untuk mengaplikasikan bahasa pemrograman python pada algoritma *Naïve Bayes* dan *Support Vector Machine* adalah *Jetbrains Pycharm* 5.0.4.

4.2.1. Pengolahan Data Menggunakan Algoritma *Naïve Bayes*



Gambar 4.2 *Output* Klasifikasi Algoritma *Naïve Bayes*

4.2.2. Pengolahan Data Menggunakan Algoritma *Support Vector Machine*



Gambar 4.3 *Output* Klasifikasi Algoritma *Support Vector Machine*

4.2.3. Hasil Akurasi Menggunakan *K-fold Cross Validation*

Setelah diolah menggunakan algoritma *Naïve Bayes* dan *Support Vector Machine*, kedua algoritma tersebut harus dihitung akurasinya menggunakan *K-fold Cross Validation*. Seperti yang sudah dijelaskan di bab sebelumnya *K-fold Cross Validation* berfungsi untuk mencari akurasi dari algoritma dengan membagi data menjadi k data dan dilakukan sebanyak k kali. Pada penelitian ini peneliti menggunaka 10 fold yang berarti data akan dibagi menjadi 10 kelompok dan dilakukan iterasi sebanyak 10 kali. Hasil dari akurasi tersebut apabila mendekati 0 maka algoritma tersebut dianggap tidak bagus atau tidak akurat, namun apabila hasil akurasi mendekati 1 atau sama dengan 1 maka akurasi dari algoritma tersebut dapat dikatakan sangat bagus atau sangat akurat.

```
iterasi ke - 1
Akurasi untuk data test kelompok ke 1 : 0.742857142857

iterasi ke - 2
Akurasi untuk data test kelompok ke 2 : 0.728571428571

iterasi ke - 3
Akurasi untuk data test kelompok ke 3 : 0.728571428571

iterasi ke - 4
Akurasi untuk data test kelompok ke 4 : 0.757142857143

iterasi ke - 5
Akurasi untuk data test kelompok ke 5 : 0.814285714286

iterasi ke - 6
Akurasi untuk data test kelompok ke 6 : 0.771428571429

iterasi ke - 7
Akurasi untuk data test kelompok ke 7 : 0.714285714286

iterasi ke - 8
Akurasi untuk data test kelompok ke 8 : 0.728571428571

iterasi ke - 9
Akurasi untuk data test kelompok ke 9 : 0.771428571429

iterasi ke - 10
Akurasi untuk data test kelompok ke 10 : 0.785714285714

rata - rata akurasi :  0.754285714286
```

Gambar 4.4 Hasil Akurasi Menggunakan Algoritma *Naïve Bayes*

Avalaiable at:
http://journal.unj.ac.id/unj/index.php/pinter/article/view/16217

```
iterasi ke - 1
Akurasi untuk data test kelompok ke 1 : 1.0

iterasi ke - 2
Akurasi untuk data test kelompok ke 2 : 1.0

iterasi ke - 3
Akurasi untuk data test kelompok ke 3 : 1.0

iterasi ke - 4
Akurasi untuk data test kelompok ke 4 : 0.985714285714

iterasi ke - 5
Akurasi untuk data test kelompok ke 5 : 0.942857142857

iterasi ke - 6
Akurasi untuk data test kelompok ke 6 : 1.0

iterasi ke - 7
Akurasi untuk data test kelompok ke 7 : 1.0

iterasi ke - 8
Akurasi untuk data test kelompok ke 8 : 1.0

iterasi ke - 9
Akurasi untuk data test kelompok ke 9 : 1.0

iterasi ke - 10
Akurasi untuk data test kelompok ke 10 : 1.0

rata - rata akurasi :  0.992857142857
```

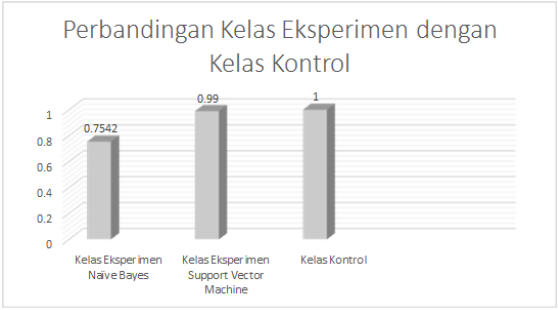
Gambar 4.5 Hasil Akurasi Menggunakan Algoritma *Support Vector Machine*

4.3. Pembahasan

Tabel 4.3 Perbandingan Hasil Akurasi Algoritma *Naïve Bayes* dan *Support Vector Machine*

Iterasi	Naïve Bayes	Support Vector Machine
1	0.7428	1.0
2	0.7285	1.0
3	0.7285	1.0
4	0.7571	0.9857
5	0.8142	0.9428
6	0.7714	1.0
7	0.7142	1.0
8	0.7285	1.0
9	0.7714	1.0
10	0.7857	1.0
Rata-rata	0.7542	0.99

Tabel 4.3 adalah tabel perbandingan hasil akurasi algoritma *Naïve Bayes* dan *Support Vector Machine* dengan menggunakan *K-fold Cross Validation*. Selanjutnya penulis membandingkan hasil rata-rata akurasi dari kelas eksperimen menggunakan dua algoritma yaitu algoritma *Naïve Bayes* dan *Support Vector Machine* dengan akurasi kelas kontrol yaitu nilai data akurat pada kondisi nyata dengan tingkat akurasi 1 atau sempurna.



Gambar 4.6 Perbandingan Kelas Eksperimen dengan Kelas Kontrol

Gambar 4.6 adalah diagram batang dari hasil akurasi kelas eksperimen dan kelas kontrol. Hasil akurasi dari kelas eksperimen yang menggunakan algoritma *Naïve Bayes* sebesar 0.7542. Untuk hasil akurasi dari kelas eksperimen yang menggunakan algoritma *Support Vector Machine* adalah sebesar 0.99, sedangkan akurasi yang seharusnya dicapai oleh kedua algoritma tersebut adalah sama dengan akurasi kelas kontrol yaitu nilai data akurat pada kondisi nyata yang seharusnya dicapai dan diberi angka 1. Namun diantara kedua algoritma tersebut yang paling mendekati akurasi kelas kontrol adalah *Support Vector Machine* dengan tingkat akurasi 0.99.

4.4. Aplikasi Hasil Penelitian

Setelah penelitian dilakukan dan mendapatkan hasil penelitian berupa satu dari dua algoritma yang lebih akurat dalam kasus seleksi kelulusan pemberkasan beasiswa BPP-PPA Fakultas Teknik Universitas Jakarta, penulis berharap penelitian ini dapat berguna untuk penelitian selanjutnya. Aplikasi dari hasil penelitian itu sendiri dapat berupa sebuah program yang dapat menyeleksi berkas secara otomatis baik yang bersifat *offline* maupun *online*, sehingga memudahkan proses seleksi berkas beasiswa tersebut. Penulis juga berharap hasil penelitian dapat digunakan untuk pembuatan *web* beasiswa dimana mahasiswa dapat mengisi dan *upload* berkas beasiswa secara *online* kemudian pada saat yang bersamaan data mahasiswa tersebut dapat diseleksi menggunakan algoritma yang telah disimpulkan pada penelitian ini.

5. Kesimpulan dan Saran

5.1 Kesimpulan

Kesimpulan yang diperoleh setelah melalui tahap-tahap perbandingan hasil akurasi algoritma *Naïve Bayes* dan *Support Vector Machine* adalah :

• Diantara algoritma *Naïve Bayes* dan *Support Vector Machine*, algoritma yang lebih akurat untuk seleksi pemberkasan Beasiswa BPP-PPA di Fakultas Teknik Universitas Negeri Jakarta adalah algoritma *Support Vector Machine*.

• Hasil klasifikasi kedua algoritma tersebut dihitung tingkat akurasinya menggunakan *K-fold Cross Validation*, dengan hasil rata-rata akurasi algoritma

Avalaiable at:
http://journal.unj.ac.id/unj/index.php/pinter/article/view/16217

Naïve Bayes adalah 0.7542 dan hasil rata-rata akurasi algoritma *Support Vector Machine* adalah 0.99.

• Baik algoritma *Naïve Bayes* maupun *Support Vector Machine* telah mampu menangani proses penyeleksian kelulusan pemberkasan beasiswa BPP-PPA Fakultas Teknik Universitas Negeri Jakarta.

• Algoritma *Support Vector Machine* menghasilkan rata-rata akurasi 0.99 dan mendekati nilai akurasi yang nyata yaitu 1, maka algoritma tersebut dinilai lebih akurat dan direkomendasikan untuk penelitian selanjutnya.

5.2 Saran

Penulis memiliki beberapa masukan untuk penelitian selanjutnya yang berhubungan dengan komputerisasi seleksi kelulusan pemberkasan Beasiswa BPP-PPA Fakultas Teknik Universitas Negeri Jakarta, dengan harapan semakin akurat dan tepat sasaran seleksi kelulusan beasiswa tersebut.

• Menambahkan *data train*, semakin banyak *data train* maka akan semakin akurat hasil dari klasifikasi.

• Memasukkan fungsi perubahan data ke bentuk vektor pada coding untuk otomatis merubah data kebentuk vektor.

Daftar Pustaka:

E, Turban dkk. (2005). *Decicion Support Systems and Intelligent Systems.* Yogyakarta:Andi Offset.

Departemen Pendidikan Nasional. (2005). *Kamus Besar Bahasa Indonesia Edisi Ketiga*. Jakarta: Balai Pustaka.

Direktorat Jendral Pendidikan Tinggi. (2015). *Pedoman Umum Beasiswa Bantuan Biaya Pendidikan Peningkatan Prestasi Akademik (PPA).* Direktorat Jendral Pembelajaran dan Kemahasiswaan Kementerian Riset, Teknologi, dan Pendidikan Tinggi.

Fakultas Teknik. 2015. *Buku Panduan Penyusunan Skripsi dan Non Skripsi*. Fakultas Teknik, Universitas Negeri Jakarta.

Han, Jiawei dkk. (2012). *Data Mining Concepts and Techniques, 3rd ed*. Waltham, USA: Elsevier Inc.

Hastuti, Khafiizh. (2012). *Analisis Komparasi Algoritma Klasifikasi Data Mining untuk Prediksi Mahasiswa Non Aktif.* Seminar Nasional Teknologi Informasi dan Komunikasi Terapan, 2(1):242-243.

Kohavi, Ron. (1995). *A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. International Joint Conference on Artifical Intellegence*, 14(2):1137-1145.

Octaviani, Pusphita Anna dkk. (2014). *Penerapan Metode Klasifikasi Support Vector Machine (SVM) pada Data Akreditasi Sekolah Dasar (SD) di Kabupatern Magelang*. Seminar Nasional Informatika, 3(4):812.

Prasetyo, Eko. (2012). *Data Mining : Konsep dan Aplikasi menggunakan MATLAB, 1st ed*. Yogyakarta:Andi Offset.

Prasetyo, Eko. (2014). *DATA MINING – Mengolah Data menjadi Informasi Menggunakan Matlab*. Yogyakarta: Andi Offset.

Avalaiable at:
http://journal.unj.ac.id/unj/index.php/pinter/article/view/16217