



Implementasi Algoritma K-Nearest Neighbor Sebagai Pendukung Keputusan Klasifikasi Penerima Beasiswa PPA dan BBM

Sumarlin

STIKOM Uyelindo Kupang

Naskah Diterima : 11 Maret 2015; Diterima Publikasi : 13 April 2015

Abstract

In line with the growth in the academic field especially college, scholarship is a problem that is interesting to study. Several studies in the field of computers for the screening or classification scholarships have been carried out in the academic authorities to minimize the error in awarding scholarships. This study discusses the classification of PPA and BBM scholarships based on variables that have been determined by applying the k-nearest neighbor algorithm. The process of selecting awardees PPA and BBM requires a decision support system (DSS) to help provide alternative solutions. The results of the classification system will be used as a decision in awarding scholarships to students who submit. Results of testing to measure the performance of k - nearest neighbor algorithm using cross validation method, Confusion Matrix and the Receiver Operating Characteristic (ROC) curve, the accuracy obtained for PPA scholarships reached 88.33% with a value of 0.925 area under curve (AUC) dataset of 227 records, while accuracy is obtained for fuel BBM scholarships reached 90% with a value of 0.937% AUC dataset of 183 records, accuracy for PPA and BBM scholarships reached 85,56% and AUC value 0,958. Because AUC values were in the range of 0.9 to 1.0 the method falls within the category of very good (excellent).

Keywords: Decision Support System; K-nearest neighbor; Classification; Scholarship

Abstrak

Sejalan dengan pertumbuhan dalam bidang akademik khususnya perguruan tinggi, beasiswa merupakan masalah yang menarik untuk diteliti. Beberapa penelitian dalam bidang komputer untuk penyeleksian atau klasifikasi beasiswa telah banyak dilakukan dalam meminimalisir adanya kekeliruan pihak akademik dalam pemberian beasiswa. Penelitian ini membahas tentang klasifikasi beasiswa peningkatan prestasi akademik (PPA) dan bantuan belajar mahasiswa (BBM) berdasarkan variabel-variabel yang telah ditentukan dengan menerapkan algoritma *k-nearest neighbor*. Proses penyeleksian penerima beasiswa peningkatan prestasi akademik dan bantuan belajar mahasiswa membutuhkan sebuah sistem pendukung keputusan (SPK) untuk membantu memberikan solusi yang alternatif. Hasil klasifikasi dari sistem ini akan digunakan sebagai keputusan dalam pemberian beasiswa bagi mahasiswa yang mengajukannya. Hasil *testing* untuk mengukur performa algoritma *k-nearest neighbor* menggunakan metode *cross validation*, *Confusion Matrix* dan *kurva Receiver Operating Characteristic(ROC)*, akurasi yang diperoleh untuk beasiswa peningkatan prestasi akademik mencapai 88,33% dengan nilai *Area Under Curva(AUC)* 0,925 dari 227 *record dataset*, sedangkan akurasi yang diperoleh untuk beasiswa Bantuan belajar mahasiswa mencapai 90% dengan nilai AUC 0,937 dari 183 *record dataset*, akurasi yang diperoleh untuk gabungan beasiswa peningkatan prestasi akademik dan bantuan belajar mahasiswa mencapai 85,56% dan nilai AUC 0,958. Karena nilai AUC berada dalam rentang 0,9 sampai 1,0 maka metode tersebut masuk dalam kategori sangat baik (*excellent*).

Kata Kunci : Sistem Pendukung Keputusan; *K-Nearest Neighbor*; Klasifikasi; Beasiswa

1. Pendahuluan

Setiap lembaga pendidikan perguruan tinggi terdapat banyak jenis beasiswa yang ditawarkan kepada mahasiswa yang unggul atau berprestasi dan kurang mampu. Untuk mengantisipasi agar beasiswa Peningkatan Prestasi Akademik (PPA) dan Bantuan Belajar Mahasiswa (BBM) tersalurkan kepada yang berhak, maka diperlukan suatu sistem untuk

mengambil keputusan agar penyalurannya tepat sasaran, tepat waktu, dan tepat jumlah.

Klasifikasi yang didasarkan pada sistem informasi merupakan teknik memetakan (mengklasifikasikan) data ke dalam satu atau beberapa kelas yang sudah didefinisikan sebelumnya. Ada banyak teknik klasifikasi yang dapat digunakan, diantaranya adalah *k-nearest neighbor*. Penggunaan algoritma yang tepat dapat meningkatkan keakuratan

*) Penulis korespondensi: sumar.lin94@yahoo.com

keputusan yang diambil. Metode klasifikasi algoritma *k-nearest neighbor* merupakan salah satu metode pengklasifikasian data yang memiliki konsistensi yang kuat, dengan cara mencari kasus dengan menghitung kedekatan antara kasus baru dengan kasus lama berdasarkan pencocokan bobot (Kusrini dan Lutfhi, 2009). Algoritma ini lebih efektif dalam melakukan training data yang besar dan dapat menghasilkan data yang lebih akurat.

Konsep penelitian dengan algoritma *K-nearest neighbor* telah banyak dilakukan oleh peneliti-peneliti sebelumnya diantaranya yaitu, *k-nearest neighbor* merupakan teknik yang sangat sederhana, efisien dan efektif dalam bidang pengenalan pola, kategori teks, pengolahan objek dan lain-lain, karena kesederhanaan pengolahannya dan mampu melakukan *training* data dalam jumlah yang besar (Bathia, 2010). Salah satu masalah dari algoritma ini adalah efek yang sama dari semua atribut yang terdapat pada data baru dan data lama dalam dataset pelatihan (Moradian dan Baraani, 2009). *K-nearest neighbor* mampu melakukan training pada dataset penyakit diabetes untuk melihat dampak negatif hilangnya nilai imputasi dan solusi untuk penyembuhan. Akurasi algoritma *k-nearest Neighbor* diatas rata-rata (Christobel *et al.*, 2013). *K-nearest neighbor* yang diterapkan pada data konsumen yang menggunakan jasa keuangan kredit kendaraan bermotor, pada penelitian ini *k-nearest neighbor* digunakan untuk penentuan resiko kredit kendaraan bermotor (Leidiyana, 2013).

K-nearest neighbor digunakan untuk memprediksi pengajuan kredit oleh debitur berdasarkan tingkat kemiripan sejumlah nilai variabel *k* (Pandie, 2012). *Nearest neighbor* tidak berkolerasi dan jarak antara sampel uji dengan sampel pelatihan tidak dihitung secara independen, serta harus mempertimbangkan antara sampel pelatihan yang berbeda (Xu, 2013). Algoritma *k-nearest neighbor* digunakan untuk mengklasifikasi pesan singkat pada twitter dan blog. Algoritma *k-nn* dapat mengurangi waktu pemrosesan dan memberikan akurasi yang baik dalam sampel pengujian (Khamar, 2013). *K-nearest neighbor* digunakan untuk klasifikasi tulisan tangan atau deteksi iris. *K-nearest neighbor* mengklasifikasikan kasus berdasarkan kesamaan data dalam sampel pengujian (Katarina, 2013).

K-nearest neighbor yang diadaptasi berdasarkan metode deteksi kesalahan diterapkan pada ruang fitur berjenis low dimensi untuk mendeteksi kesalahan potensial. Pada penelitian ini *Knn* dikombinasikan dengan DM (*diffusion map based*) untuk memantau proses manufaktur semikonduktor (Li dan Zang, 2014). Data *mining* merupakan proses analisis dan eksplorasi. Peta klasifikasi data menjadi kelompok-kelompok atau kelas yang telah ditetapkan. Pada penelitian ini digunakan algoritma *k-nearest neighbor classifier* untuk melakukan klasifikasi data pemasaran langsung. Untuk mengukur akurasi dari

algoritma *k-nearest neighbor*, digunakan metode *cross validation* yang melibatkan estimasi akurasi dengan baik. Hasil dari penelitian ini menunjukkan bahwa hasil akurasi klasifikasi dan prediksi data pemasaran langsung dengan algoritma *k-nearest neighbor* relatif tinggi (Govindrajan dan Chandrasekaran, 2010).

Permasalahan yang sering timbul adalah banyaknya calon penerima beasiswa PPA dan BBM yang tidak tepat sasaran, sehingga diperlukan sebuah sistem yang dapat mengklasifikasikan calon penerima beasiswa PPA dan BBM berdasarkan data training yang diambil dari data mahasiswa penerima beasiswa PPA dan BBM sebelumnya (*dataset*). Sehingga pihak perguruan tinggi bisa mengatasi sejak dini permasalahan tersebut. Penggunaan teknik data mining dengan algoritma *k-nearest neighbor* diharapkan mampu memberikan informasi yang berguna tentang teknik klasifikasi calon penerima beasiswa PPA dan BBM.

Tujuan dari penelitian ini adalah mengimplementasikan Algoritma *k-nearest neighbor* sebagai pendukung keputusan klasifikasi mahasiswa calon penerima beasiswa. Membangun aplikasi sistem pendukung keputusan yang mampu mengklasifikasikan mahasiswa calon penerima beasiswa dengan menggunakan algoritma *k-nearest neighbor*.

2. Kerangka Teori

2.1 Sistem Pendukung Keputusan

Sistem Pendukung Keputusan merupakan Sistem berbasis komputer yang interaktif, yang dapat menyelesaikan masalah-masalah yang tak terstruktur (Subakti, 2002). Sistem pendukung keputusan menggunakan data, memberikan antarmuka pengguna yang mudah, dan dapat menggabungkan pemikiran pengambil keputusan (Turban, 2005). Sistem pendukung keputusan juga disebut sebagai suatu pendekatan dalam mengambil sebuah keputusan.

Adapun komponen-komponen dalam sistem pendukung keputusan adalah sebagai berikut (Xu *et al.*, 2013).

1. *Data Management*. Termasuk *database*, yang mengandung data yang relevan untuk berbagai situasi dan diatur oleh *software* yang disebut *Database Management Systems* (DBMS).
2. *Model Management*. Melibatkan model finansial, statistik, *management science*, atau berbagai model kuantitatif lainnya, sehingga dapat memberikan ke sistem suatu kemampuan analitis, dan manajemen *software* yang diperlukan.
3. *Communication (dialog subsystem)*. *User* dapat berkomunikasi dan memberikan perintah pada SPK melalui subsistem ini. Ini berarti menyediakan antarmuka.

4. *Knowledge Management*. Subsystem optional ini dapat mendukung subsystem lain atau bertindak sebagai komponen yang berdiri sendiri.

2.2 Proses Pengambilan Keputusan

Pengambilan keputusan di dalam suatu organisasi merupakan hasil suatu proses komunikasi dan partisipasi yang terus menerus dari keseluruhan organisasi. Hasil keputusan tersebut dapat merupakan pernyataan yang disetujui antar alternatif atau antar prosedur untuk mencapai tujuan tertentu. Pendekatannya dapat dilakukan melalui pendekatan yang bersifat individual atau kelompok, sentralisasi atau desentralisasi, partisipasi atau tidak berpartisipasi maupun demokratis atau konsensus. Persoalan pengambilan keputusan, pada dasarnya adalah bentuk pemilihan dari berbagai alternatif tindakan yang mungkin dipilih, yang prosesnya melalui mekanisme tertentu, dengan harapan akan menghasilkan sebuah keputusan yang terbaik.

2.3 Metode Klasifikasi

Klasifikasi merupakan proses penemuan model (fungsi) yang menggambarkan dan membedakan kelas data atau konsep yang bertujuan agar bisa digunakan untuk memprediksi kelas dari objek yang label kelasnya tidak diketahui [(Han dan Kamber, 2006). Klasifikasi merupakan bagian dari data mining, dimana Data mining merupakan suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan didalam database. Data mining juga proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstrasi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar (Vercellis, 2009).

Proses klasifikasi didasarkan pada empat komponen (Gorunescu, 2011).

1. *Kelas*
Variabel dependen yang berupa kategorikal yang merepresentasikan 'label' yang terdapat pada objek. Contohnya: resiko penyakit jantung, resiko kredit, *customer loyalty*, jenis gempa.
2. *Predictor*
Variabel independen yang direpresentasikan oleh karakteristik (atribut) data. Contohnya: merokok, minum alkohol, tekanan darah, tabungan, aset, gaji.
3. *Training dataset*
Satu set data yang berisi nilai dari kedua komponen di atas yang digunakan untuk menentukan kelas yang cocok berdasarkan *predictor*.
4. *Testing dataset*
Berisi data baru yang akan diklasifikasikan oleh model yang telah dibuat dan akurasi klasifikasi dievaluasi.

2.4 Algoritma K-Nearest Neighbor

K-nearest neighbor merupakan sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut. Data pembelajaran diproyeksikan ke ruang berdimensi banyak, dimana masing-masing dimensi merepresentasikan fitur dari data. Ruang ini dibagi menjadi bagian-bagian berdasarkan klasifikasi data pembelajaran. Nilai k yang terbaik untuk algoritma ini tergantung pada data, secara umumnya, nilai k yang tinggi akan mengurangi efek noise pada klasifikasi, tetapi membuat batasan antarasetiap klasifikasi menjadi lebih kabur.

Ada banyak cara untuk mengukur jarak kedekatan antara data baru dengan data lama (data *training*), diantaranya *euclidean distance* dan *manhattan distance (city block distance)*, yang paling sering digunakan adalah *euclidean distance* (Bramer, 2007), yaitu:

$$\sqrt{(a_1 - b_1)^2 + (a_2 + b_2)^2 + \dots + (a_n - b_n)^2} \quad (1)$$

Untuk melakukan pembobotan variabel dilakukan dengan metode perbandingan pasangan, bobot variabel ditentukan dengan cara normalisasi vektor eigen, yang diasosiasikan dengan nilai eigen maksimum pada suatu matriks rasio sebelum membandingkan harus ditentukan skala nilai pengaruh atau penting antara variabel, seperti yang terlihat pada Tabel 1 (Saaty, 1980).

Tabel 1. Skala untuk perbandingan pasangan

Nilai	Defenisi
1	Sama Penting
2	Sama hingga cukup penting
3	Cukup penting
4	Cukup penting hingga tinggi kepentingannya
5	Tinggi kepentingannya
6	Tinggi kepentingannya hingga sangat tinggi
7	Sangat tinggi kepentingannya
8	Kepentingannya sangat tinggi hingga amat sangat tinggi
9	Kepentingannya amat sangat tinggi

Untuk melakukan pembobotan nilai variabel yang mempunyai kemiripan digunakan fungsi sebagai berikut (Larose, 2005).

$$\text{Different}(a_i, b_i) \begin{cases} 0 & \text{jika } a_i = b_i \\ 1 & \text{jika } a_i \neq b_i \end{cases} \quad (2)$$

Dimana a_i dan b_i adalah nilai kategori. Jika nilai variabel antara dua *record* yang dibandingkan sama maka nilai jaraknya 0, jika nilai variabel mempunyai kemiripan atau kedekatan maka diberi bobot 0,5, sebaliknya jika nilai variabel berbeda maka akan diberi bobot 1 artinya tidak mirip sama sekali.

Misalkan variabel status dengan nilai menikah dan menikah, maka nilai kedekatannya 0, jika menikah dan janda maka nilai kedekatannya antara 0,5, sedangkan jika nilai menikah dan belum menikah maka nilai kedekatannya adalah 1.

Tabel 2. Skala Pembobotan tingkat kemiripan variabel

Nilai (%)	Defenisi	Bobot
0	Jika $a_i = b_i$ (nilai data sama)	0
50	Jika (a_i) mirip dengan (b_i)	0,5
100	Jika $a_i \neq b_i$ (nilai tidak sama)	1

Untuk menghitung kemiripan kasus, digunakan rumus (KUsrini dan Ludhfi, 2009).

$$Similarity(p, q) = \frac{\sum_{i=1}^n f(p_i, q_i) * w_i}{w_i} \quad (3)$$

Keterangan :

- p = Kasus baru
- q = Kasus yang ada dalam penyimpanan
- n = Jumlah atribut dalam tiap kasus
- p_i = Kasus baru dalam Atribut individu antara 1 sampai dengan n
- q_i = Kasus yang ada dalam penyimpanan Atribut individu antara 1 sampai dengan n
- f = Fungsi *similarity* atribut i antara kasus p dan kasus q
- w_i = Bobot yang diberikan pada atribut ke-i kedalam Atribut individu antara 1 sampai dengan n

2.5 Standarisasi Data

Dalam proses standarisasi, data ditransformasi ke dalam rentang nilai [-1..1] atau [0..1]. Dalam penelitian ini digunakan standarisasi nilai dengan rentang [0..1]. Normalisasi bisa dilakukan dengan *min-max normalization* atau *Z-score standardization*. Jika data *training* terdiri dari atribut campuran antara numerik dan kategori, lebih baik gunakan *min-max normalization* (Larose, 2005).

$$V_n = \frac{v - \min_A}{\max_A - \min_A} (new_{\max_A} - new_{\min_A}) + new_{\min_A} \quad (4)$$

Keterangan :

V_n = Nilai Variabel ke-n

Dari rumus di atas \min_A dan \max_A merupakan nilai-nilai *minimum* dan *maximum* dari atribut A, *min max normalization* memetakan sebuah nilai v dari A menjadi V_n dalam rentang $[new_min_A, new_max_A]$.

2.6. Metode Evaluasi dan Validasi Algoritma Data Mining

a. Cross Validation

Cross validation merupakan pengujian standar yang dilakukan untuk memprediksi *error*. Data

training dibagi secara random ke dalam beberapa bagian dengan perbandingan yang sama kemudian *error* dihitung bagian demi bagian, selanjutnya hitung rata-rata seluruh *error* untuk mendapatkan *error rate* secara keseluruhan (Kohavi, 1995).

Evaluasi digunakan untuk mengukur kinerja metode klasifikasi, dalam penelitian ini digunakan untuk mengukur keakuratan metode klasifikasi yang diukur dengan akurasi, *precision* dan *recall*. *Recall* didefinisikan sebagai persentase antara data kelas data buruk yang dikelaskan dengan benar dan data kelas data buruk yang salah diprediksi ke kelas data baik. *Precision* adalah persentase dari kelas data buruk yang dikelaskan dengan benar dan kelas yang seharusnya termasuk kelas data baik tetapi dikelaskan sebagai kelas data buruk.

b. Confusion Matrix

Metode ini menggunakan tabel matriks, jika data set hanya terdiri dari dua kelas, kelas yang satu dianggap sebagai positif dan yang lainnya negatif (Bramer, 2007).

Tabel 3. Model *Confusion Matrix*

Klasifikasi yang benar	Diklasifikasi sebagai	
	+	-
+	<i>True positive</i>	<i>False negative</i>
-	<i>False positive</i>	<i>True negative</i>

Keterangan tabel :

1. *True positives* merupakan jumlah *record* positif yang diklasifikasikan sebagai positif,
2. *False positives* merupakan jumlah *record* negatif yang diklasifikasikan sebagai positif,
3. *False negatives* merupakan jumlah *record* positif yang diklasifikasikan sebagai negatif,
4. *True negatives* merupakan jumlah *record* negatif yang diklasifikasikan sebagai *negative*.

c. Kurva ROC

Kurva ROC menunjukkan akurasi dan membandingkan klasifikasi secara visual. ROC mengekspresikan *confusion matrix*. ROC adalah grafik dua dimensi dengan *false positives* sebagai garis horisontal dan *true positives* sebagai garis vertikal. *The area under curve* (AUC) dihitung untuk mengukur perbedaan performansi metode yang digunakan. AUC dihitung menggunakan rumus (Liao, 2007):

$$\theta^a = \frac{1}{mn} \sum_{j=1}^n \sum_{i=1}^m \psi(x_i^r, x_j^r) \quad (5)$$

Dimana,

$$K = \begin{cases} 1 & Y < X \\ \frac{1}{2} & Y = X \\ 0 & Y > X \end{cases} \quad (6)$$

Keterangan :

K = Jumlah algoritma yang dikomparasi

X = Output Positif

Y = Output Negatif

2.6 Beasiswa PPA dan BBM

Tiap-tiap warga negara berhak mendapatkan pengajaran. Hak setiap warga negara tersebut telah dicantumkan dalam Pasal 31 (1) Undang-Undang Dasar 1945. Berdasarkan pasal tersebut, maka Pemerintah dan pemerintah daerah wajib memberikan layanan dan kemudahan, serta menjamin terselenggaranya pendidikan yang bermutu bagi setiap warga negara tanpa diskriminasi, dan masyarakat berkewajiban memberikan dukungan sumber daya dalam penyelenggaraan pendidikan. Untuk menyelenggarakan pendidikan yang bermutu diperlukan biaya yang cukup besar. Oleh karena itu bagi setiap peserta didik pada setiap satuan pendidikan berhak mendapatkan biaya pendidikan bagi mereka yang orang tuanya tidak mampu membiayai pendidikannya, dan berhak mendapatkan beasiswa bagi mereka yang berprestasi (Santoso, 2010).

Beasiswa Peningkatan Prestasi Akademik (PPA) merupakan beasiswa yang diberikan untuk peningkatan pemerataan dan kesempatan belajar bagi mahasiswa yang mengalami kesulitan membayar biaya pendidikannya sebagai akibat krisis ekonomi, terutama bagi mahasiswa yang berprestasi akademik. Beasiswa Bantuan Belajar Mahasiswa (BBM) merupakan beasiswa yang berupa bantuan yang diberikan kepada mahasiswa yang mengalami kesulitan membayar biaya pendidikannya. Sama dengan PPA, tujuannya membantu meringankan beban orang tua dari kalangan ekonomi lemah.

3. Metodologi

3.1 Bahan dan Alat Penelitian

Bahan penelitian berupa data sekunder yang diperoleh dari laporan data penerima beasiswa PPA dan BBM. Data yang berhubungan dengan penelitian ini antara lain yaitu *history* data mahasiswa penerima beasiswa PPA dan BBM sebagai data *training* dan data mahasiswa calon penerima beasiswa PPA dan BBM sebagai data sampel uji (*testing*). Pengumpulan data dalam penelitian ini melalui studi pustaka.

Alat yang digunakan dalam penelitian ini adalah Perangkat Keras berupa Laptop Compaq dengan spesifikasi prosesor intel Centrino 2 Ghz, Memory minimal 1 Gb dan Harddisk minimal 50 Gb. Sedangkan perangkat lunak yang digunakan dalam penelitian ini adalah *Borland Delphi 7* sebagai bahasa pemrogramannya dan *database* yang digunakan adalah *MySql*.

3.2 Prosedur Penelitian

Tahapan dalam prosedur ini dapat dijelaskan sebagai berikut

a. Identifikasi dan Perumusan Masalah

Pada tahap ini dilakukan identifikasi parameter atau atribut-atribut apa saja yang dapat digunakan untuk mengukur penentuan klasifikasi mahasiswa calon penerima beasiswa PPA dan BBM. Tahapan ini juga dapat menentukan model analisis yang akan digunakan dalam penentuan klasifikasi calon penerima beasiswa baik secara kuantitatif maupun kualitatif untuk sistem pendukung keputusan. Perumusan masalah dibuat setelah melakukan identifikasi masalah.

b. Tujuan Penelitian

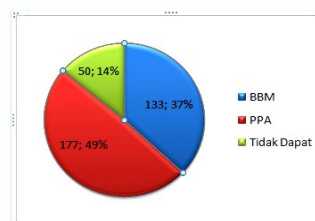
Tahapan ini digunakan untuk mengembangkan sistem informasi yang mampu untuk mendukung pengambilan keputusan dalam mengklasifikasi mahasiswa calon penerima beasiswa PPA dan BBM dengan menggunakan algoritma *k-nearest neighbor*.

c. Studi Kepustakaan

Pada tahap ini dilakukan proses pengumpulan literatur seperti, jurnal internasional maupun nasional yang sesuai dengan topik penelitian, juga berupa artikel yang relevan dengan topik penelitian.

d. Studi Lapangan

Studi lapangan dilakukan di Kampus STIKOM Uyelindo Kupang dengan cara melakukan observasi dan wawancara dengan narasumber (bagian kemahasiswaan). Studi lapangan dilakukan untuk memperoleh data-data yang berkaitan dengan tema penelitian yaitu berupa data mahasiswa penerima beasiswa PPA dan BBM. Dari hasil observasi ditemukan beberapa data penerima beasiswa PPA dan BBM pada tahun ajaran 2012/2013. Mahasiswa penerima beasiswa Peningkatan Prestasi Akademik (PPA) sebanyak 177 mahasiswa, penerima Bantuan Belajar Mahasiswa sebanyak 133 mahasiswa dan sebanyak 60 mahasiswa yang tidak lolos seleksi beasiswa PPA dan BBM dari berbagai program studi. Terdapat ketidakseimbangan data dengan jumlah data penerima beasiswa yang termasuk ke dalam kategori PPA mendominasi sebesar 49%, BBM 37% dan tidak dapat 14% dari keseluruhan data.



Gambar 1. Presentasi Data Penelitian

e. Pengumpulan Data

Pada tahap ini dilakukan pengumpulan data yang diperoleh dari studi lapangan. Adapun Data yang

digunakan dalam penelitian ini yaitu berupa data *history* mahasiswa penerima beasiswa PPA dan BBM yang digunakan sebagai data latih (*data training*) untuk mengklasifikasi mahasiswa calon penerima beasiswa PPA dan BBM. Juga digunakan data mahasiswa calon penerima beasiswa PPA dan BBM sebagai sampel untuk menguji sistem yang akan dibangun dengan analisis algoritma *k-nearest Neighbor*. Dari data yang dikumpulkan diperoleh rata-rata IPK mahasiswa penerima beasiswa PPA dan BBM berkisar dari 3,00 – 3,50 dan didominasi oleh mahasiswa semester 6, serta rata-rata penerima beasiswa PPA dan BBM yang penghasilan orangtuanya lebih dari satu juta rupiah

f. Penentuan Variabel

Pada tahap ini akan ditentukan variabel/parameter apa saja yang digunakan dalam pengolahan data pada sistem yang akan dibangun dengan algoritma *k-nearest Neighbor*. Variabel-variabel ini akan diperoleh dari data *history* mahasiswa penerima beasiswa PPA dan BBM yaitu berupa : Indeks Prestasi Kumulatif (IPK), semester, piagam penghargaan, tagihan listrik, penghasilan orang tua dan jumlah tanggungan orang tua. Penentuan variabel dilakukan dengan memilih atribut-atribut yang berpengaruh dalam proses penyeleksian beasiswa PPA dan BBM, pemberian nilai parameter/variabel disesuaikan dengan data mentah (*data training*). Atribut/variabel dan nilai variabel dapat dilihat pada Tabel 4.

Tabel 4. Penentuan Variabel dan Nilai Variabel

No.	Variabel	Nilai Variabel
1.	Indeks Prestasi (IPK)	1. 2,50 – 3,00
		2. 3,01 – 3,50
		3. 3,51 – 4,00
2.	Semester	1. Semester II
		2. Semester III
		3. Semester IV
		4. Semester V
		5. Semester VI
		6. Semester VII
3.	Penghasilan orang tua	1. Rp 100.000 – 500.000
		2. Rp 500.001 – 1.000.000
		3. Rp 1.000001 -2.000.000
		4. > 2.000.000
4.	Piagam	1. Tidak ada
		2. Wilayah / Propinsi
		3. Nasional
		4. Internasional
5.	Tagihan Listrik	1. Rp 10.000 – 50.000
		2. Rp 50.001 – 150.000
		3. Rp > 150.000
6.	Tanggungan orang tua	1. 1 orang
		2. 2 orang
		3. 3 orang
		4. > 3 orang

g. Penentuan Bobot Variabel

Untuk melakukan pengukuran jarak antara atribut, maka akan dilakukan pembobotan pada atribut. Bobot jarak atribut diberikan nilai antara 0 sampai 1, pembobotan variabel dilakukan dengan metode perbandingan pasangan variabel dengan skala yang telah ditentukan pada Tabel 5 dan 6.

Tabel 5. Pembobotan Variabel Beasiswa PPA

No.	Nama Variabel	Bobot
1	Indek Prestasi Kumulatif (IPK)	0,4118
2	Semester	0,2292
3	Piagam	0,1496
4	Penghasilan Orang Tua	0,0969
5	Tanggungan Orang Tua	0,0684
6	Tagihan Listrik	0,0440

Tabel 6. Pembobotan Variabel Beasiswa BBM

No.	Nama Variabel	Bobot
1	Indek Prestasi Kumulatif (IPK)	0,3642
2	Semester	0,2630
3	Piagam	0,1603
4	Penghasilan Orang Tua	0,0980
5	Tanggungan Orang Tua	0,0694
6	Tagihan Listrik	0,0451

h. Penentuan Nilai Kedekatan Variabel

Menentukan nilai kedekatan, biasanya berada pada nilai antara 0 s/d 1. Nilai 0 artinya kedua kasus sama, nilai 0,5 jika memiliki nilai yang mirip, sebaliknya untuk nilai 1 data tidak mirip. Untuk menentukan nilai kedekatan antara variabel dapat digunakan persamaan (2), jika nilai variabel mempunyai kemiripan seperti yang terlihat pada tabel 7, dan persamaan (4) jika memiliki rentang nilai pengukuran yang jauh berbeda seperti yang terlihat pada Tabel 8.

Tabel 7. Nilai Variabel IPK

Atribut	Nilai Atribut 1	Nilai atribut 2	Bobot
	2,50 – 3,00	2,50 – 3,00	0
Indeks	2,50 – 3,00	3,01 – 3,50	1
Prestasi	2,50 – 3,00	3,51 – 4,00	1
Kumulatif (IPK)	3,01 – 3,50	3,01 – 3,50	0
	3,01 – 3,50	3,51 – 4,00	0,5
	3,51 – 4,00	3,51 – 4,00	0

Tabel 8. Nilai Variabel Penghasilan orang tua

Atribut	Nilai Atribut 1	Nilai atribut 2	Bobot
	100.000 – 500.000	100.000 – 500.000	0,63
	100.000 – 500.000	500.000 – 1 juta	0,44
Penghas ilan	100.000 – 500.000	1.000.001 – 2 juta	0,47
	100.000 – 500.000	> 2.000.0000	0,66
Orang Tua	500.000 – 1.000.000	500.000 – 1 juta	0,50
	500.000 – 1.000.000	1.000.001 – 2 juta	0,33
	500.000 – 1.000.000	> 2.000.0000	0,60
	1.000.001 – 2 juta	1.000.001 – 2 juta	0,50
	1.000.001 – 2 juta	> 2.000.0000	0,50
	> 2.000.0000	> 2.000.0000	0,75

3.3 Perangan Sistem

Perancangan Sistem adalah menentukan bagaimana mencapai sasaran yang ditetapkan yang melibatkan pembentukan (*configuring*) perangkat lunak dan komponen perangkat keras sistem dimana setelah pemasangan sistem akan memenuhi spesifikasi yang dibuat pada akhir fase analisis sistem (Scott, 2001). Berdasarkan kebutuhan informasi dan kebutuhan data yang ada pada analisis sebelumnya, maka berikutnya akan dilakukan perancangan implementasi algoritma *k-nearest neighbor* dalam pengambilan keputusan klasifikasi penerima beasiswa PPA dan BBM dengan menggunakan beberapa *tool's* untuk pemodelan proses maupun perancangan basis data. Perancangan yang digunakan dalam penelitian ini adalah *Unified Modeling Language* (UML) dengan tahapan mengidentifikasi pelaku bisnis, membuat proses bisnis, Diagram model *use case*, diagram aktifitas, diagram *sequence*, *class diagram* dan *Data Flow Diagram* (DFD).

3.4 Analisa Keputusan

Hasil akan diputuskan dengan sistem pengambil keputusan dengan menggunakan algoritma *k-nearest neighbor* untuk melakukan klasifikasi mahasiswa penerima beasiswa PPA dan BBM. Untuk pemilihan variabel-variabel yang akan dianalisa diambil dari data mahasiswa yang mengajukan beasiswa.

3.5 Desain Perancangan Form Aplikasi

Pada tahap ini akan dibuat perancangan *user interface* yang dibuat untuk merancang form aplikasi, yang meliputi : menu *login*, *input data*, proses perhitungan dan menu hasil perhitungan (*dashboard*).

4. Hasil dan Pembahasan

Penelitian ini menggunakan 320 data *training* yang terdiri dari 177 data penerima beasiswa PPA, 133 data penerima beasiswa BBM dan 50 data mahasiswa yang tidak menerima beasiswa PPA dan BBM. Untuk mengetahui hasil analisis dari algoritma *k-nearest neighbor* maka dilakukan perhitungan manual dengan menggunakan data sampel seperti berikut :

Data *training* :

1. Ipk = 3,59, semester = 6, piagam = tidak ada, penghasilan orang tua = 2 juta, tanggungan = 3 orang dan tagihan listrik = 100 ribu (PPA).
2. Ipk = 3,19, semester = 7, piagam = tidak ada, penghasilan orang tua = 1.800.000, tanggungan = 2 orang dan tagihan listrik = 57 ribu (BBM).
3. Ipk = 2,90, semester = 2, piagam = nasional, penghasilan orang tua = 1.500.000, tanggungan = 2 orang dan tagihan listrik = 100 ribu (BBM).

Misalkan ada kasus baru sebagai data *testing* akan dihitung kedekatannya dengan kasus lama (data *training*)

Data *testing* :

Ipk = 3,40, semester = 3, piagam = tidak ada, penghasilan orang tua = 1.200.000, tanggungan = 3 orang dan tagihan listrik = 50 ribu.

Untuk menghitung kedekatan kasus antara data *training* dan data *testing* di atas, digunakan persamaan (3). Untuk mengetahui apakah data testing termasuk klasifikasi beasiswa PPA atau BBM, dapat dilakukan langkah-langkah sebagai berikut (Pandie, 2012):

1. Menghitung kedekatan antara data baru dengan data nomor 1.

Diketahui :

- a. Kedekatan nilai variabel IPK (3,59 dan 3,40) = 0,50
- b. Bobot variabel IPK = 0,41
- c. Kedekatan nilai variabel semester (6 dan 3) = 1
- d. Bobot variabel semester = 0,10
- e. Kedekatan nilai variabel piagam (tidak ada dengan tidak ada) = 0
- f. Bobot variabel piagam = 0,07
- g. Kedekatan variabel penghasilan orang tua (2.000.000 dengan 1.200.000) = 0,50.
- h. Bobot variabel penghasilan orang tua = 0,23
- i. Kedekatan variabel tanggungan orang tua (3 dan 3) = 0.
- j. Bobot variabel tanggungan orang tua = 0,15
- k. Kedekatan variabel tagihan listrik (100.000 dengan 50.000) = 0,50
- l. Bobot variabel tagihan listrik = 0,04.

Dihitung :

$$Jarak = \frac{(a * b) + (c * d) + (e * f) + (g * h) + (i * j) + (k * l)}{b + d + f + h + j + l}$$

$$Jarak = \frac{(0,5 * 0,41) + (1 * 0,10) + (0 * 0,07) + (0,50 * 0,23) + (0 * 0,15) + (0,50 * 0,04)}{0,41 + 0,10 + 0,07 + 0,23 + 0,15 + 0,04}$$

$$Jarak = 0,4400$$

2. Menghitung kedekatan antara data baru dengan data nomor 2.

Diketahui :

- a. Kedekatan nilai variabel IPK (3,19 dan 3,40) = 0
- b. Bobot variabel IPK = 0,36
- c. Kedekatan nilai variabel semester (7 dan 3) = 1
- d. Bobot variabel semester = 0,10
- e. Kedekatan nilai variabel piagam (tidak ada dengan tidak ada) = 0
- f. Bobot variabel piagam = 0,07
- g. Kedekatan variabel penghasilan orang tua (500.000 dengan 1.200.000) = 0,33
- h. Bobot variabel penghasilan orang tua = 0,26
- i. Kedekatan variabel tanggungan orang tua (4 dan 3) = 0,50.

- j. Bobot variabel tanggungan orang tua = 0,16
- k. Kedekatan variabel tagihan listrik (30.000 dengan 50.000) = 0,38
- l. Bobot variabel tagihan listrik = 0,05

Dihitung :

Jarak = 0,2848

- 3. Menghitung kedekatan antara kasus baru dengan kasus nomor 3.

Diketahui :

- a. Kedekatan nilai variabel IPK (2,90 dan 3,40) = 1
- b. Bobot variabel IPK = 0,36
- c. Kedekatan nilai variabel semester (2 dan 3) = 0,5
- d. Bobot variabel semester = 0,10
- e. Kedekatan nilai variabel piagam (nasional dengan tidak ada) = 1
- f. Bobot variabel piagam = 0,07
- g. Kedekatan variabel penghasilan orang tua (1.500.000 dengan 1.200.000) = 0,50
- h. Bobot variabel penghasilan orang tua = 0,26
- i. Kedekatan variabel tanggungan orang tua (2 dan 3) = 0,50
- j. Bobot variabel tanggungan orang tua = 0,16
- k. Kedekatan variabel tagihan listrik (100.000 dengan 50.000) = 0,50
- l. Bobot variabel tagihan listrik = 0,05.

Dihitung :

Jarak = 0,7150

- 4. Memilih kasus dengan kedekatan terdekat. Dari langkah 1, 2 dan 3 dapat diketahui bahwa nilai terendah adalah kasus nomor 2. Berarti kasus yang terdekat dengan kasus baru adalah kasus nomor 2 pada data *training*.
- 5. Menggunakan klasifikasi dari kasus dengan kedekatan terdekat. Berdasarkan hasil yang diperoleh pada langkah 3, maka klasifikasi dari 3 kasus yang akan digunakan untuk mengklasifikasi kasus baru. Yaitu kemungkinan calon penerima beasiswa akan mendapat beasiswa BBM.

4.1. Verifikasi Hasil Perhitungan Ms. Excel dan Sistem

Untuk mengetahui hasil dengan menggunakan ms. Excel dan sistem / aplikasi yang telah dibuat digunakan 360 data sampel dan diuji validitas dengan data *testing* yang sama yaitu Ipk = 3,03, semester = 3 , piagam = tidak ada, penghasilan orang tua = 500.000, tanggungan = 2 orang dan tagihan listrik = 40 ribu. Maka diperoleh hasil yang sama dari perhitungan dengan menggunakan ms. Excel dan sistem yaitu klasifikasi beasiswa PPA dengan jarak kedekatan 0,1334.

4.2. Pengujian Algoritma

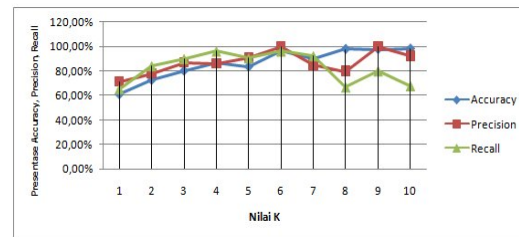
- a. Cross Validation

Dalam penelitian ini digunakan *10 fold-cross validation* dimana 360 *record* pada data *training* dibagi secara random ke dalam 10 bagian dengan

perbandingan yang sama kemudian *error rate* dihitung bagian demi bagian, selanjutnya hitung rata-rata seluruh *error rate* untuk mendapatkan *error rate* secara keseluruhan. Data diuji tingkat erornya menggunakan teknik cross validation dengan *10 fold* dan terdapat 35 *recordfold* ke-1 sampai ke-9 dan 45 *record* pada *fold* ke-10. Hasil pengujian menunjukkan persentase tingkat *precision* data terendah berada pada K = 1 yaitu 71,43% dan akurasi masing-masing mencapai 60,83% dan *recall* 65,22%.

Tabel 9. Akurasi, *Precision*, *Recall* algoritma *k-nn*

	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>
K=1	60,83%	71,43%	65,22%
K=2	72,50%	77,78%	84%
K=3	80%	87,10%	90%
K=4	86,67%	86,21%	96,15%
K=5	83,33%	90,62%	90,62%
K=6	96,67%	100%	96,30%
K=7	90,00%	85,00%	92,33%
K=8	97,7%	80%	66,67%
K=9	97,50%	100%	80,00%
K=10	98,00%	92,33%	67,77%



Gambar 2. Hasil proses 10 fold cross validation

- b. Confusion Matrix

Perhitungan kedekatan kasus lama pada data *training* dengan kasus baru pada data *testing*, diketahui dari 360 *record* data, 288 diklasifikasikan Beasiswa, 22 data diklasifikasibeasiswa tetapi ternyata tidak, 20 data *class* tidakdiklasifikasi sesuai, dan 30 data diklasifikasi tidakternyata beasiswa. Tingkat akurasi penerapan algoritma *k-nearest neighbor* ini sebesar 85.56%, *precision* 90,57% dan *recall* 92,90%.

Tabel 10. Confusion matrix untuk *k-nearest neighbor*

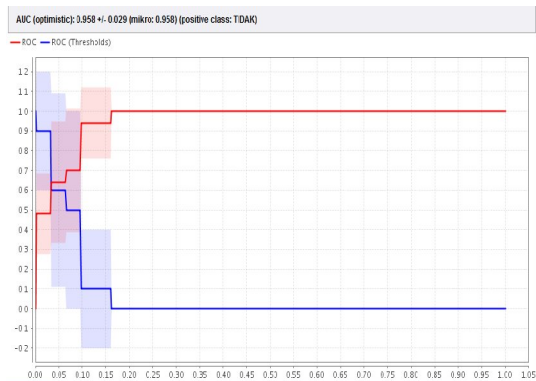
accuracy: 85.56% +/- 4.00% (mikro: 85.56%)			
	true BEASISWA	true TIDAK	class precision
pred BEASISWA	288	30	90.57%
pred TIDAK	22	20	47.82%
class recall	92.90%	40.00%	

- c. Kurva ROC

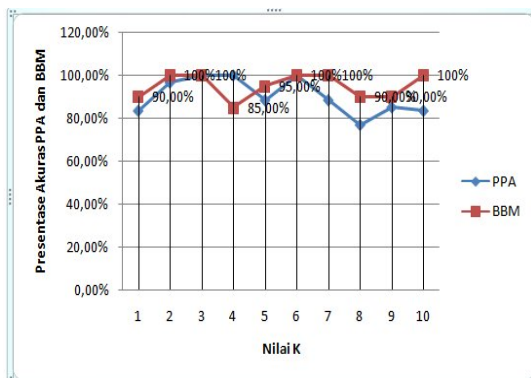
Kurva ROC menunjukkan akurasi dan membandingkan klasifikasi secara visual. ROC mengekspresikan *confusion matrix*. ROC adalah grafik dua dimensi dengan *false positives* sebagai

garis horizontal dan *true positive* sebagai garis vertical (Xu *et al.*, 2013). Hasil perhitungan divisualisasikan dengan kurva ROC.

Terdapat dua titik garis pada kurva ROC yaitu garis *true negative rate specificity* dan garis *true negative rate sensitivity*. *Specificity* didefinisikan sebagai kemampuan alat tes untuk mendeteksi data negatif sedangkan *sensitivity* didefinisikan sebagai kemampuan alat tes untuk mendeteksi data positif. Garis *specificity* juga disebut sebagai batas atas normal dan garis *sensitivity* disebut sebagai garis batas bawah normal, tingkat kehandalan diperoleh dengan menghitung luas area dibawah garis batas normal atas kurva. Berdasarkan pengelompokan di atas maka dapat disimpulkan bahwa metode *k-nearest neighbor* termasuk klasifikasi sangat baik untuk mengklasifikasi beasiswa PPA karena memiliki nilai AUC (*Area Under Curva*) antara 0.90-1.00 yaitu 0,925.

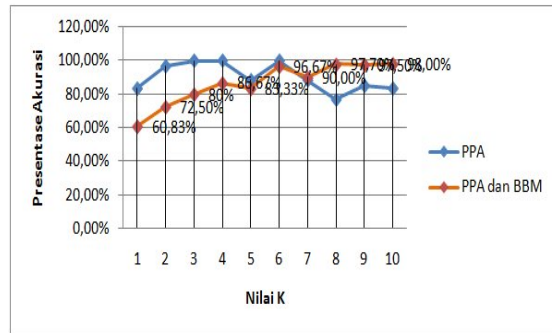


Gambar 3. Kurva ROC metode *k-nearest neighbor*



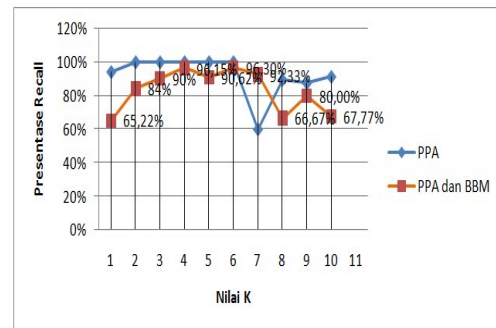
Gambar 4. Perbandingan akurasi beasiswa PPA dan BBM

Berdasarkan gambar 4 diperoleh nilai presentasi akurasi terendah berada pada K=8 untuk klasifikasi beasiswa PPA yaitu 76,67% dan mencapai kestabilan data pada K=1 sampai K=7 dengan presentasi akurasi rata-rata di atas 80%. Nilai K optimal diperoleh pada K=3 dan K=6.



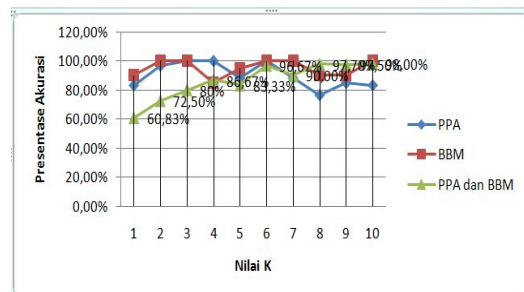
Gambar 5. Perbandingan akurasi beasiswa PPA dan Gabungan

Berdasarkan gambar 5 diperoleh nilai presentasi akurasi terendah berada pada K=1 untuk klasifikasi gabungan beasiswa PPA dan BBM yaitu 60,83% dan mencapai kestabilan data pada K=3 sampai K=7, dan K=9 sampai K=10 dengan presentasi akurasi rata-rata di atas 80%. Nilai K optimal diperoleh pada K=7.



Gambar 6. Perbandingan akurasi beasiswa BBM dan Gabungan

Berdasarkan gambar 6 diperoleh nilai presentasi akurasi terendah berada pada K=1 untuk klasifikasi gabungan beasiswa PPA dan BBM yaitu 60,83% dan mencapai kestabilan data pada K=2 sampai K=10 dengan presentasi akurasi rata-rata di atas 80%. Nilai K optimal diperoleh pada K=4 dan K=10



Gambar 7. Perbandingan akurasi beasiswa PPA, BBM dan Gabungan

Berdasarkan gambar 7 diperoleh nilai presentasi akurasi terendah berada pada K=1 klasifikasi gabungan beasiswa PPA dan BBM yaitu 60,83% dan mencapai kestabilan akurasi pada K=3 sampai K=7

dengan presentasi akurasi di atas 80%. K optimal diperoleh pada K=6.

5. Kesimpulan

Penelitian ini dilakukan dengan mengimplementasikan algoritma *k-nearest neighbor* pada data penerimaan beasiswa PPA dan BBM. Untuk mencari data yang berkualitas, dilakukan preprosesing sebelum diterapkan ke dalam algoritma. Kedekatan antara kasus baru dengan kasus lama dilakukan untuk menentukan pada kelas mana kasus baru akan diklasifikasikan. Dengan membangun sistem pendukung keputusan untuk mengklasifikasikan beasiswa PPA dan BBM yang ditentukan atau dianalisis dengan menggunakan algoritma *k-nearest neighbor*. Ada enam variabel yang digunakan yaitu indeks prestasi kumulatif, semester, piagam penghargaan, tanggungan orang tua, pendapatan orang tua dan tagihan listrik.

Hasil uji verifikasi yang ada menunjukkan bahwa sistem pendukung keputusan yang dibuat dengan menggunakan analisis algoritma *k-nearest neighbor* menghasilkan keluaran yang sama dengan perhitungan manual yang dilakukan dengan microsoft excel, dimana keluaran sistem berupa nilai terendah yang dijadikan kasus terdekat untuk mengklasifikasi kasus baru sama dengan hasil perhitungan microsoft excel.

Untuk mengukur kinerja algoritma *k-nearest neighbor* digunakan metode *Cross Validation*, *Confusion Matrix* dan Kurva ROC, dalam penelitian ini menggunakan *10-fold cross validation* untuk klasifikasi beasiswa PPA dan BBM. Dari *227 record dataset* beasiswa PPA diperoleh tingkat *accuracy* mencapai 77,96% dan termasuk klasifikasi sangat baik karena memiliki nilai AUC berada diantara 0,90-1,00, yaitu 0,925, sedangkan untuk beasiswa BBM digunakan *183 record dataset* dan diperoleh tingkat *accuracy* mencapai 97,28% dengan nilai AUC 0,937, termasuk dalam klasifikasi sangat baik. Akurasi. Untuk gabungan beasiswa PPA dan BBM diperoleh *accuracy* mencapai 85,56% dan memiliki nilai AUC 0,958 juga termasuk klasifikasi sangat baik.

Daftar Pustaka

- Bhatia, M., Vandana., 2010. Survey of nearest neighbor techniques. *International Journal of Computer Science and Information Security* 8, 1947-5500.
- Bramer, M., 2007. *Principles of Data Mining*, Springer, London.
- Christobel, A.C., Sivaprakasam, P., 2013. A New Classwise k Nearest Neighbor (CKNN) method for the classification of diabetes dataset. *International Journal of Engineering and Advanced Technoloi* (2), 2249 – 8958.
- Gorunescu, F., 2011. *Data Mining : Concepts, Models, and Techniques*. Verlag Berlin Heidelberg, Springer.
- Govindrajan, M., Chandrasekaran, R.M., 2010. Evaluation of k-Nearest Neighbor classifier performance for direct marketing. *Expert Systems with Applications* 37, 253–258
- Han, J., Kamber, M., 2006. *Data Mining Concept and Tehniques*, San Fransisco, Morgan Kauffman.
- Katarina, A., Singh, M.D., 2013. A review of data classification using K-Nearest Neighbour algorithm. *International Journal of Emerging Technology and Advanced Engineering* 3, 2250-2459.
- Khamar, K., 2013. Short text classification using knn based ondistance function. *International Journal of Advanced Research*.
- Kohavi, R., 1995. A study of cross validation and bootstrap for accuracy estimation and model selection. *Proceedings of the International JointConference on Articial Intelligence*, vol. 2, 1137-1143.
- Kusrini dan Lutfhi, E.T., 2009. *Algoritma Data Mining*, Andi Publishing, Yogyakarta.
- Moradian, M., Baraani, A., 2009. K-Nearest Neighbor Based Association Algorithm. *Journal of Theoretical and Applied Information Technology* 6, 123 – 129.
- Larose, D.T., 2005. *Discovering Knowledge in Data*. New Jersey, John Willey & Sons, Inc
- Leidiyana, H., 2013. PenerapanAlgoritma *K-Nearest Neighbor* untuk penentuan resiko kredit kepemilikan kendaraan bermotor. *Jurnal Penelitian Ilmu Komputer, System Embedded & Logic* 1(1), 65-76.
- Liao., 2007. *Recent Advances in Data Mining of Enterprise Data Algoritms and Application*. World Scientific Publishing, Singapore.
- Li, Y., Zhang, X., 2014. Diffusion maps based k-nearest-neighbor rule techniquefor semiconductor manufacturing process fault detection, *Chemometrics and Intelligent Laboratory Systems* 136, 47-57
- Pandie, E.S.Y., 2012. Implementasi Algoritma Data Mining K-Nearest Neighbour(K-Nn) DalamPengambilan Keputusan Pengajuan Kredit, *Seminar Nasional Sains dan Teknik*, vol. 4, Kupang, November 13, 31-34
- Santoso, D., 2010. *Pedoman Beasiswa Peningkatan Prestasi Akademik (PPA) dan Bantuan Belajar Mahasiswa (BBM)*, Jakarta.
- Saaty, T.L., 1980. *The Analytic Hierarchy Process*, McGraw-Hill, New York
- Scott, G. M., 2001. *Prinsip-prinsip Sistem Informasi Manajemen*. Mc.Graw-Hill, Jakarta.
- Surbakti, I., 2002. *Sistem Pendukung Keputusan*, Institut Teknologi Sepuluh November, Surabaya.
- Turban, E., 2005. *Decicion Support System and intelligent System*, Andi Publisher. Yogyakarta

Journal of Innovative Computing, Information
and Control 9, 1349-4198.

- Vercellis., C., 2009. Business Intelligent Data Mining and Optimization for Decision Making. Southern Gate, Chichester, West Sussex.
- Xu, Y., Zhu, Q., Chen, Y., Pan, J.S., 2013. An Improvement To The Nearest Neighbor Classifier And Face Recognition Experiments. International