

Penentuan Jurusan Sekolah Menengah Atas Menggunakan Metode *K-Nearest Neighbor Classifier* Pada SMAN 16 Semarang

Ari Sulistiyo¹

^{1,3}Jurusan Teknik Informatika, FASILKOM UDINUS
Jln. Nakula I No.5-11 Semarang 50131 INDONESIA
¹111201106240@mhs.dinus.ac.id

Abstract--Majors in high school is a grouping of learners to a group of subjects aimed at adjusting the abilities and interests of students in the chosen field so as to increase the interest and the achievement of learners, and the system used majors that still use manual calculation method that is less effective because of would require quite a long time because the number of students who will be majors. This research will discuss about the placement of students with majors classify based on grades, test scores of national, interests and recommendations of BK to the classification system using K-Nearest Neighbor Classifier is a method of classification which the shortest distance calculation used as a measure to classify a new case based on the size similarity of old cases that have been classified before. By using the 96 training data and test data 64 obtained the highest results at 51 of data classified correctly and the remaining 13 are classified with less precise. From these results match the level of accuracy obtained in class results actually majors against the majors prediction classes using K-Nearest Neighbor Classifier on the value of $k = 7$ amounted to 79.68% and included the classification is very good because it has the AUC values between 0.90 - 1:00, which amounted to 0962.

Index Terms—Majors, Senior High School, Classifier, K-Nearest Neighbor

I. PENDAHULUAN

Penjurusan siswa yang dilakukan pada Sekolah Menengah Atas sesuai minat dan dilakukan sejak peserta didik mendaftar ke SMA/MA memiliki tujuan untuk memberikan kesempatan kepada peserta didik mengembangkan kompetensi sikap, kompetensi keterampilan peserta didik sesuai dengan minat, bakat, dan kemampuan akademik dalam sekelompok mata pelajaran keilmuan [1]. Penentuan jurusan akan berdampak pada jenjang akademik berikutnya dan akan mempengaruhi bidang ilmu atau studi bagi siswa-siswi yang akan melanjutkan ke jenjang perguruan tinggi nantinya sehingga penjurusan yang tidak tepat bisa sangat merugikan siswa dan masa depannya.

Penempatan siswa sesuai dengan kapasitas kemampuannya atau sering disebut dengan penjurusan siswa di sekolah menengah ditentukan oleh kemampuan akademik yang didukung oleh faktor minat, karena karakteristik suatu ilmu menuntut karakteristik yang sama dari yang mempelajarinya. Dengan demikian, siswa yang mempelajari suatu ilmu yang sesuai dengan karakteristik kepribadiannya akan merasa senang ketika mempelajari ilmu tersebut. Minat dapat mempengaruhi kualitas pencapaian hasil belajar siswa dalam bidang studi tertentu. Seorang siswa yang berminat pada Matematika misalnya, akan memusatkan perhatiannya lebih banyak ke bidang Matematika daripada siswa lain. Karena pemusatan

perhatian intensif terhadap materi, siswa akan belajar lebih giat dan mencapai prestasi yang diinginkan [2].

Pengambilan keputusan penjurusan sesuai kurikulum 2013 yaitu menentukan jurusan saat siswa duduk dibangku kelas X dilakukan oleh pihak sekolah dengan melihat beberapa faktor diantaranya yaitu nilai Rapor SMP/MTs atau yang sederajat, nilai Ujian Nasional SMP/MTs atau yang sederajat, rekomendasi guru Bimbingan dan Konseling/Konselor di SMP/MTs atau yang sederajat. Penjurusan

yang tersedia di SMA meliputi Ilmu Alam (IPA), Ilmu Sosial (IPS), Ilmu Bahasa [1]. Disamping itu, penjurusan juga diselenggarakan untuk menyesuaikan kemampuan dan minat peserta didik terhadap bidang yang dipilihnya. Penempatan penjurusan yang sesuai akan meningkatkan minat dan memberikan kenyamanan seseorang dalam belajar. Dengan dasar kemampuan yang sama diharapkan dalam kegiatan pembelajaran dapat berjalan dengan lancar tanpa ada yang mengalami kesulitan dan dapat meningkatkan minat serta prestasi belajar peserta didik. Sebaliknya, kurangnya minat untuk belajar akibat kesalahan dalam memilih jurusan. [2]

Akan tetapi karena terlalu banyaknya siswa yang akan dijuruskan membuat guru BK merasa kesulitan untuk melakukan suatu perhitungan untuk menentukan jurusan tersebut karena proses penjurusan masih dilakukan dengan cara manual dengan mengelompokan nilai-nilai yang ada berdasarkan jurusan yang sesuai, dan masih mempertimbangkan Minat dan Rekomendasi dari guru BK dari sekolah sebelumnya, sehingga cara tersebut kurang efektif dan

efisien karena harus bekerja dua kali dan membutuhkan waktu yang cukup lama.

Klasifikasi merupakan proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data, dengan tujuan untuk dapat memperkirakan kelas dari suatu objek yang labelnya tidak diketahui. Model itu sendiri bisa berupa aturan “jika-maka”, berupa decision tree, formula matematis atau neural network. Metode-metode klasifikasi antara lain C4.5, RainForest, Naïve Bayesian, neural network, genetic algorithm, fuzzy, case-based reasoning, dan k-Nearest Neighbor. Metode Klasifikasi dengan Algoritma K-Nearest Neighbor (k-NN) adalah suatu metode yang menggunakan algoritma supervised, dimana hasil dari sampel uji yang baru diklasifikasikan berdasarkan mayoritas dari kategori pada k-NN yang bertujuan untuk mengklasifikasi objek baru berdasarkan atribut dan training samples [3].

Oleh karena itu Penelitian ini akan menganalisis penerapan teknik klasifikasi K-Nearest Neighbor untuk mengelompokkan siswa Sekolah Menengah Atas (SMA) dalam Penjurusan Minat Program Studi.

II. STUDI PUSTAKA

2.1 Penelitian Terkait

Penulis memulai penelitian ini dengan terlebih dahulu melakukan studi kepustakaan dari penelitian-penelitian dan sumber-sumber lain. Dari studi kepustakaan itu, penulis menemukan beberapa penelitian yang mendorong untuk mengangkat tema seperti diatas. Penelitian tersebut membahas tentang topik yang terkait dengan penelitian penulis, antara lain adalah penelitian mengenai algoritma yang digunakan penulis yang akan diangkat oleh penulis.

1. Bahar dengan judul “Penentuan Jurusan Sekolah Menengah Atas dengan Algoritma Fuzzy C-Means”
2. Sedy Winanta, Yetli Oslan, dan Gunawan Santoso dengan judul “Implementasi Metode Bayesian dalam Penjurusan di SMA Bruderan Purworejo”
3. Henny Leidyana dengan judul “Penerapan Algoritma K-Nearest Neighbor Untuk Penentuan Resiko Kredit Kepemilikan Kendaraan Bemotor”

2.2 Tinjauan Pustaka

A. Data Mining

Data Mining merupakan proses pencarian pola dan relasi-relasi yang tersembunyi dalam sejumlah data yang besar dengan tujuan untuk melakukan klasifikasi, estimasi, prediksi, association rule, clustering, deskripsi dan visualisasi [6]

B. Klasifikasi

Klasifikasi adalah tugas pembelajaran sebuah fungsi target f yang memetakan setiap himpunan atribut x ke salah satu label class y yang telah didefinisikan sebelumnya . Klasifikasi dapat juga diartikan suatu proses untuk menemukan suatu model atau fungsi yang menggambarkan dan membedakan kelas data atau konsep dengan tujuan

dapat menggunakan model untuk memprediksi kelas objek yang label class-nya tidak diketahui [6].

C. K-Nearest Neighbor (K-NN)

K-Nearest Neighbor (K-NN) termasuk kelompok instance-based learning. Algoritma ini juga merupakan salah satu teknik lazy learning. kNN dilakukan dengan mencari kelompok k objek dalam data training yang paling dekat (mirip) dengan objek pada data baru atau data testing [6].

D. Euclidian Distance

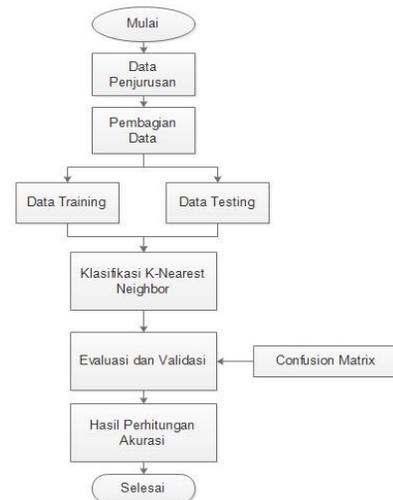
Metode K-Nearest Neighbor dilakukan perhitungan kuadrat jarak euclid (query instance) masing-masing obyek terhadap data sampel yang diberikan. Dalam perhitungan jarak ada beberapa rumus jarak contohnya seperti Euclidian Distance, Square euclidean distance, dan Manhattan distance. Karena secara umum untuk mendefinisikan jarak antara dua objek x dan y , maka digunakan rumus jarak Euclidean Distance seperti terlihat pada persamaan (1) [7].

$$D(a, b) = \sqrt{\sum_{k=1}^d (X_k - Y_k)^2}$$

$$\begin{aligned} X^1 &= [X_1, X_2, \dots, X_p] \text{ dan } Y^1 = [Y_1, Y_2, \dots, Y_p] \\ D^2_{(x,y)} &= (X_1 - Y_1)^2 + (X_2 - Y_2)^2 + \dots + (X_p - Y_p)^2 \\ &= (X - Y) (X - Y) = (Y - X) (Y - X) \end{aligned}$$

dimana matriks $D(a,b)$ adalah jarak skalar dari kedua vektor a dan b dari matriks dengan ukuran d dimensi.

2.3 Kerangka Pemikiran



Gambar 1 Kerangka Pemikiran

III. METODE PENELITIAN

3.1 Instrumen Penelitian

Dalam membangun Penelitian ini digunakan bahasa pemrograman PHP (Hypertext Processor), sehingga dibutuhkan beberapa perangkat lunak yang mendukungnya yaitu sebagai

berikut :

- a. Web Browser (Mozilla Firefox atau Google Chrome)
Web Browser digunakan untuk mencari studi literate tentang penelitian sejenis, dan untuk menjalankan program yang akan dibuat nantinya.
- b. Web Server (XAMPP).
XAMPP digunakan sebagai PHP server pada saat implementasi metode. Sedangkan web browser digunakan untuk menjalankan simulasi atau sebagai media interface.
- c. Editor (Notepad++)
Editor yang digunakan untuk membuat simulasi program
- d. Microsoft Office Word
Digunakan untuk membantu melakukan dokumentasi penelitian yang telah dibuat.
- e. Weka
Digunakan untuk membantu melakukan proses perhitungan akurasi dalam klasifikasi.

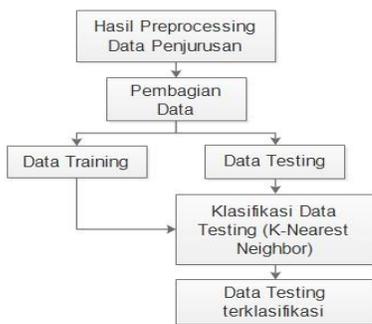
Sedangkan Perangkat Keras yang digunakan memiliki spesifikasi sebagai berikut :

- a. Sebuah laptop type ASUS N43SL dengan processor Intel Core i3-2330 2.20GHz, Memory 4 GB, Operating System Windows 8.1 64Bit
- b. Printer HP 1515, yang digunakan untuk mencetak laporan yang telah dibuat.

3.2 Pengumpulan Data

Data yang digunakan dalam penelitian ini adalah data siswa hasil penjurusan pada SMAN 16 Semarang pada tahun 2014 sebanyak 160 siswa. Dataset berisi beberapa variable diantaranya Nilai Rapor (B. Indonesia, B. Inggris, Matematika, IPA, IPS) Nilai Ujian Nasional (B. Indonesia, B. Inggris, Matematika, IPA, IPS) Rata-rata nilai IPA, IPS, Bahasa, Minat, Rekomendasi BK, dan Hasil Penjurusan.

3.3 Teknik Analisis Data



Gambar 2 Teknik Analisis Data

Hasil dari proses *preprocessing* akan dibagi menjadi dua bagian yaitu Data Training (Latih) dan Data Testing (Uji) dengan prosentase 60% : 40%. Selanjutnya akan dilakukan perhitungan terhadap data uji dengan menggunakan metode *K-Nearest Neighbor* untuk menghasilkan probabilitas setiap kelas dengan menghitung bobot kedekatannya dengan data latih yang ada. Nilai probabilitas setiap kelas yang dihasilkan

digunakan untuk melakukan klasifikasi data uji sehingga menghasilkan data uji yang terklasifikasi.

IV. HASIL & PEMBAHASAN

Didalam bab ini akan dibahas mengenai data yang akan digunakan dalam penelitian, data tersebut akan di hitung menggunakan algoritma *K-Nearest Neighbor Classifier* dan kemudian diuji menggunakan Confusion Matrix dan Kurva ROC.

4.1 Data yang digunakan

Data yang digunakan dalam penelitian ini adalah sebanyak 160 data hasil penjurusan siswa dengan atribut Nilai Rapor, Nilai Ujian, Minat Penjurusan, Penjurusan Rekomendasi Guru BK, dan Hasil Penjurusan. Dari 160 data penjurusan siswa yang ada akan dibagi menjadi 2 bagian, yaitu data latih dan data data uji yang masing-masing dibagi menjadi 96 data latih dan 64 data uji.

Tabel 1 Sampel Data Latih

No	Nilai Rapor					Nilai Ujian					Minat	Rekom BK	Hasil
	B. Indo	B. Ing	Mtk	IPA	IPS	B. Indo	B. Ing	Mtk	IPA	IPS			
1	8.4	7.4	7	7.1	6.7	8	7.4	7.25	8.25	7.4	IPA	IPA	IPA
2	8.6	6.6	8.1	7	7.6	7.6	7	7.5	7.25	8.2	IPS	Bahasa	IPS
3	7.4	6.6	7.6	7.8	7.4	7.8	8	7.75	7.75	8	IPS	IPS	IPS
4	7.5	8.2	6.6	7.1	7.7	8.2	7.4	7	8	7.2	IPA	IPS	IPA
5	7.8	8.3	6.5	7.6	8	7.8	7.2	7.25	8.5	7.7	IPA	IPA	IPA
6	7.5	6.6	7.9	7.8	6.5	8.4	7.8	8	7.5	7.5	IPS	IPS	IPA
7	8.4	7.8	6.9	6.7	7.8	7.2	8.6	7	7.75	6.8	Bahasa	Bahasa	Bahasa
8	7.7	7.5	6.8	8	7.6	7.4	7.6	6.5	7.25	8	IPS	Bahasa	Bahasa
9	7.3	8.3	7.3	7.4	8.1	8	7.8	7.25	7	7.1	IPA	IPS	IPS
10	7.3	8.1	7.2	7.2	6.5	7	8	7	7	6.8	Bahasa	IPS	IPS

4.2 Perhitungan Algoritama K-Nearest Neighbor

Tahapan awal pada proses K-Nearest Neighbor yaitu dengan menentukan hubungan jarak bobot antar atribut Minat dan Rekomendasi pada data yang ada sehingga dapat dilakukan perhitungan. Pada tabel berikut menunjukkan hubungan bobot antar atribut

Tabel 2 Hubungan bobot antar atribut

	IPA	IPS	Bahasa
IPA	0	1	1
IPS	1	0	1
Bahasa	1	1	0

Sebagai contoh akan dilakukan perhitungan dengan data pada tabel berikut:

Tabel 3 Data Uji

No	Nilai Rapor					Nilai Ujian					Minat	Rekom BK	Hasil
	B. Indo	B. Ing	Mtk	IPA	IPS	B. Indo	B. Ing	Mtk	IPA	IPS			
1	7.5	8.5	8.3	7.7	7.2	7.4	7.6	7.25	7	7.8	IPS	Bahasa	?

Kemudian data uji akan dihitung jarak kedekatannya dengan data latih pada Tabel 1 dengan menggunakan persamaan berikut:

$$D(a, b) = \sqrt{\sum_{k=1}^d (X_k - Y_k)^2}$$

Sehingga didapatkan hasil sebagai berikut:

Tabel 4 Data Latih beserta jarak setelah diurutkan

No	Nilai Rapor					Nilai Ujian					Minat	Rekom BK	Hasil	Jarak (D-)
	B. Indo	B. Ing	Mtk	IPA	IPS	B. Indo	B. Ing	Mtk	IPA	IPS				
8	7.7	7.5	6.8	8	7.6	7.4	7.6	6.5	7.25	8	IPS	Bahasa	Bahasa	2.050610
9	7.3	8.3	7.3	7.4	8.1	8	7.8	7.25	7	7.1	IPA	IPS	IPS	2.206808
10	7.3	8.1	7.2	7.2	6.5	7	8	7	7	6.8	Bahasa	IPS	IPS	2.352127
2	8.6	6.6	8.1	7	7.6	7.6	7	7.5	7.25	8.2	IPS	Bahasa	IPS	2.488976
3	7.4	6.6	7.6	7.8	7.4	7.8	8	7.75	7.75	8	IPS	IPS	IPS	2.516446
4	7.5	8.2	6.6	7.1	7.7	8.2	7.4	7	8	7.2	IPA	IPS	IPA	2.773536
6	7.5	6.6	7.9	7.8	6.5	8.4	7.8	8	7.5	7.5	IPS	IPS	IPA	2.865746
7	8.4	7.8	6.9	6.7	7.8	7.2	8.6	7	7.75	6.8	Bahasa	Bahasa	Bahasa	2.878368
1	8.4	7.4	7	7.1	6.7	8	7.4	7.25	8.25	7.4	IPA	IPA	IPA	2.905598
5	7.8	8.3	6.5	7.6	8	7.8	7.2	7.25	8.5	7.7	IPA	IPA	IPA	2.932576

Berdasarkan hasil perhitungan jarak diatas, jika : Nilai K ditetapkan = 5, maka didapatkan hasil :

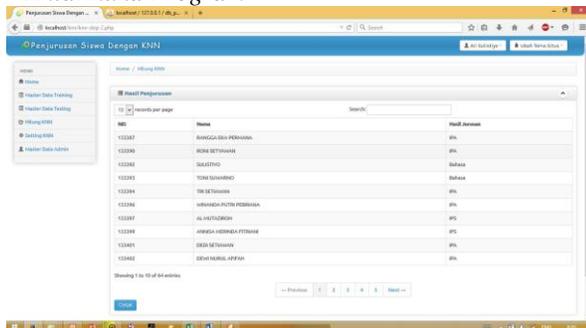
IPA = 0

IPS = 4

Bahasa = 1

Didapatkan Hasil “ IPS “ memiliki jumlah paling banyak, jadi kesimpulannya adalah siswa yang memiliki nilai sesuai data uji yang telah dihitung diprediksi masuk kedalam jurusan “ IPS “.

4.3 Antarmuka Program



Gambar 3 Tampilan Data Uji hasil klasifikasi KNN

4.4 Pengukuran Akurasi

Tahap akhir dalam penelitian ini adalah menentukan akurasi metode *K-Nearest Neighbor Classifier* terhadap data uji yang telah diklasifikasikan. Untuk mendapatkan tingkat akurasi yang baik dilakukan proses percobaan nilai k = 3, 5, 7, 9 dan 11 kemudian dibandingkan mana yang memiliki tingkat akurasi yang lebih baik.

a. Confusion Matrix

Nilai K	Prediksi Benar	Prediksi Salah	Akurasi
3	50	14	78,12%
5	49	15	76,56%
7	51	13	79,68%
9	50	14	78,12%
11	50	14	78,12%

Tabel 5 Hasil Akurasi Confusion Matrix

Jadi kesimpulan yang diperoleh yaitu, dari 160 data penjurusan yang kemudian dibagi menjadi 64 data uji dan diklasifikasikan dengan 96 data latih yang ada dilakukan percobaan menggunakan nilai k yang berbeda-beda maka nilai akurasi yang didapat berbeda-beda pula, tetapi memiliki nilai akurasi yang hampir sama sehingga diperoleh tingkat akurasi kecocokan tertinggi pada hasil penjurusan sebenarnya terhadap hasil penjurusan dari prediksi dengan menggunakan metode *K-Nearest Neighbor* didapatkan pada nilai k = 7, sebesar 79,68%.

b. Kurva ROC

Hasil perhitungan divisualisasikan dengan Kurva ROC, untuk perhitungannya digunakan Weka dan hasilnya sebagai berikut :

Terlihat pada gambar ROC Area terdapat nilai *Wighted Average* sebesar 0.962 yang disebut nilai *AOC (Area UnderCover)*.

Correctly Classified Instances	51	79.6875 %					
Incorrectly Classified Instances	13	20.3125 %					
Kappa statistic	0.6916						
Mean absolute error	0.2153						
Root mean squared error	0.3044						
Relative absolute error	48.4444 %						
Root relative squared error	64.9938 %						
Coverage of cases (0.95 level)	100 %						
Mean rel. region size (0.95 level)	71.3542 %						
Total Number of Instances	64						
=== Detailed Accuracy By Class ===							
	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.87	0.073	0.87	0.87	0.87	0.962	IPA
	0.7	0.029	0.955	0.7	0.808	0.969	IPS
	0.909	0.17	0.526	0.909	0.667	0.944	Bahasa
Weighted Avg.	0.797	0.069	0.85	0.797	0.806	0.962	

Gambar 4 Uji Akurasi dengan Weka

Untuk klasifikasi data mining, nilai AUC dapat dibagi menjadi beberapa kelompok (Gorunescu, 2011). [10]

- 0.90-1.00 = klasifikasi sangat baik
- 0.80-0.90 = klasifikasi baik
- 0.70-0.80 = klasifikasi cukup
- 0.60-0.70 = klasifikasi buruk
- 0.50-0.60 = klasifikasi salah

Berdasarkan pengelompokan diatas maka dapat disimpulkan bahwa metode kNN termasuk klasifikasi sangat baik karena memiliki nilai AUC antara 0.90 – 1.00.

V. KESIMPULAN DAN SARAN

a. Kesimpulan

Dari hasil penelitian Penentuan Jurusan Sekolah Menengah Atas Menggunakan Metode *K-Nearest Neighbor Classifier* dapat disimpulkan bahwa metode *K-Nearest Neighbor Classifier* mempunyai akurasi kecocokan dari hasil penjurusan yang sebenarnya terhadap prediksi penjurusan dengan menggunakan metode kNN dalam melakukan pengelompokan siswa kedalam jurusannya yaitu didapatkan pada nilai k = 7 yang terbesar yaitu sebesar 79,68% dan termasuk klasifikasi sangat baik karena memiliki nilai AUC antara 0.90 – 1.00, yaitu sebesar 0.962.

b. Saran

Beberapa hal yang perlu dikembangkan dalam penelitian selanjutnya dengan ruang lingkup yang sama adalah :

1. Menggunakan Data Latih yang lebih banyak dan lebih memiliki atribut yang lebih beragam untuk bisa meningkatkan nilai akurasi.
2. Menggunakan metode perhitungan klasifikasi lain, seperti *Naïve Bayes*, *C.45*, *K-Means*, dan *Support Vector Machine (SVM)* untuk mencari metode yang paling efektif dan nilai akurasi yang tinggi dalam melakukan proses penjurusan pada Sekolah Menengah Atas.

UCAPAN TERIMAKASIH

Penulis mengucapkan terima kasih kepada Universitas Dian Nuswantoro, Rektor UDINUS, Dekan Fakultas Ilmu Komputer, Kaprogdi Teknik Informatika-S1, Dosen pembimbing, serta teman-teman dan sahabat yang selama ini telah mendampingi penulis selama kuliah di Universitas Dian Nuswantoro.

REFERENSI

- [1]. M. Nur, "Peminatan pada Pendidikan Menengah," Jakarta : Menteri Pendidikan dan Kebudayaan Republik Indonesia, 2014.
- [2]. Bahar, "Penentuan Jurusan Sekolah Menengah Atas dengan Algoritma Fuzzy C-Means," Thesis Program Pasca Sarjana Udinus, Semarang, 2011.
- [3]. Larose and L. D, *Discovering Knowledge in Data*, USA: John Wiley's and Son, 2005.
- [4]. S. Winanta, Y. Oslan and G. Santoso, "Implementasi Metode Bayesian dalam Penjurusan di SMA Brudera Purworejo," EKSIS (Universitas Kristen Duta Wacana Yogyakarta), vol. 06 No. 2, pp. 21-28, 2013.
- [5]. H. Leidiyana, "Penenerapan Algoritma K-Nearest Neighbor untuk Penentuan Resiko Kredit Kepemilikan Kendaraan Bermotor," Pikel(STMIK Nusa Mandiri Jakarta), vol. 1, pp. 65-76, 2013.
- [6]. J. Han, M. Kamber and J. Pei, *Data Mining: Concepts and Techniques*, San Francisco: Morgan, 2011.
- [7]. Ian H Witten, *Data Mining Principle Machine Learning Tools and Techniques*, Burlington: Elsevier Inc, 2011.
- [8]. Bramer, *Principles of Data Mining*, London: Springer, 2007.
- [9]. Kusriani and L. ET, *Algoritma Data Mining*, Yogyakarta: Andi Publishing, 2009.
- [10]. F. Gorunescu, *Data Mining: Concepts, Models, and Techniques.*, Verlag Berlin Heidelberg : Springer, 2011.