PERBANDINGAN ALGORITMA C5.0 DAN NEAREST NEIGHBOR UNTUK MENENTUKAN KELAYAKAN PENGAJUAN KREDIT NASABAH BANK

Selvia Lorena BR Ginting[1], Bryan Philips[2]
[1,2] Program Strudi Sistem Komputer Universitas Komputer Indonesia Bandung
[1] philipus.bryanrahail@gmail.com

ABSTRAK

Data Mining merupakan teknik menambang data yang berjumlah sangat besar menjadi pola-pola tertentu. C5.0 dan Nearest Neighbor adalah algoritma-algoritma yang bisa digunakan pada metode Klasifikasi yang ada pada teknologi ini. Cara kerja dari kedua algoritma ini juga sangat menarik, karena memiliki teknik yang berbeda dalam mengklasifikasikan data. C5.0 akan mengubah data menjadi pola-pola yang berbentuk pohon keputusan, sedangkan Nearest Neighbor akan mengklasifikasi data menurut jarak kedekatan antara data yang sudah ada dengan data yang akan diuji. Karena cara kerja kedua algoritma yang berbeda, mendasari penulis untuk membandingkan kedua algoritma ini dengan menggunakan data kelayakan pengajuan kredit nasabah bank yang berada di Maluku Utara, sekaligus dapat menentukan kelayakan nasabah dalam mendapatkan kredit. Karena mengetahui jumlah data yang dimiliki bank sangat banyak, salah satunya yaitu data kelayakan nasabah dalam pengajuan kredit. Hal ini dapat menyebabkan kecerobohan dalam menentukan nasabah mana yang layak mendapatkan kredit. Sehingga dengan adanya aplikasi yang akan dibuat ini, dapat menyelesaikan masalah tersebut. Hasil akhir dari penelitian ini adalah sebuah aplikasi yang dapat membandingkan dua algoritma yaitu C5.0 dan Nearest Neighbor dalam keunggulannya mengklasifikasi data, dan juga dapat menentukan kelayakan pengajuan kredit nasabah bank.

Kata Kunci: Data Mining, C5.0, Nearest Neighbor, Klasifikasi, Kredit.

ABSTRACT

Data Mining is a technique that mines very large data into certain patterns. C5.0 and Nearest Neighbor are the algorithms that can be used in the existing Classification method of this technology. The workings of these two algorithms are also very interesting, because it has different techniques in classifying data. C5.0 will turn data into patterns in the form of decision trees, while Nearest Neighbor will classify the data according to the proximity distance between existing data and the data to be tested. Because of the way the two algorithms work differently, underlies the authors to compare these two algorithms by using the bank credit application credit feasibility data located in North Maluku, as well as to determine the eligibility of customers in obtaining credit. Because knowing the amount of data owned by the bank very much, one of which is the customer's eligibility data in the filing of credit. This can lead to carelessness in determining which customers are eligible for credit. So with the application to be made this, can solve the problem. The final result of this research is an application that can compare two algorithms that is C5.0 and Nearest Neighbor in its superiority to classify the data, and also can determine the feasibility of credit application of bank customer.

***Keywords** : Data Mining, C5.0, Nearest Neighbor, Classification, Credit*

I. PENDAHULUAN

Data mining merupakan sebuah teknik untuk menemukan pola dari data yang berjumlah sangat besar. Salah satu metode yang digunakan adalah Klasifikasi, Nearest Neighbor dan C5.0 merupakan contoh dari algoritma yang digunakan pada metode klasifikasi. Melihat bahwa dua algoritma ini mempunyai caranya sendiri untuk mengklasifikasi data, mendasari peneliti untuk membandingkan kedua algoritma ini dengan menggunakan data kelayakan pengajuan kredit nasabah bank yang ada di Maluku Utara.

Menurut definisinya Bank adalah lembaga usaha yang berfungsi untuk menyimpan dan meminjamkan uang kepada masyarakat. Kredit merupakan salah satu program bank dimana nasabah boleh meminjam uang di bank dengan syarat-syarat tertentu. Dengan

jumlah data nasabah pengajuan kredit yang begitu besar, sehingga dapat terjadinya sebuah kecerobohan dalam menentukan kelayakan kredit nasabah bank.

Dari masalah tersebut, ditemukan solusi yaitu dengan menggunakan sebuah aplikasi yang dapat menentukan kelayakan nasabah dalam pengajuan kredit dengan menggunakan dua algoritma yaitu C5.0 dan Nearest Neighbor sekaligus membandingkan kedua algoritma ini.

Aplikasi ini akan dibangun dengan menggunakan Teknologi *Data Mining* yang di integrasikan dengan algoritma *C5.0* dan dibandingkan dengan algoritma *Nearest Neighbor* untuk mengetahui algoritma mana yang lebih baik dalam mengklasifikasikan data.

II. TEORI PENUNJANG

A. Bank

Bank merupakan badan usaha yang berfungsi menyimpan uang dan meminjamkan pada nasabah[1].

B. Kredit

Kredit berasa dari kata percaya, Kredit dalam pengertian ekonomi merupakan suatu penangguhan pembayaran, ialah uang atau barang yang diterima sekarang akan dikembalikan pada masa yang akan datang [2].

C. Data Mining

Data mining merupakan teknologi untuk menambang data yang berjumlah sangat besar menjadi pola-pola tertentu.[3].

D. C5.0

Algoritma C5.0 merupakan penyempurnaan dari algoritma terdahulu yang dibentuk oleh Ross Quinlan pada tahun 1987, yaitu algoritma ID3 dan C4.5. Perbedaan utama C5.0 dari ID3 adalah[4].

- C5.0 mampu menangani atribut berkelanjutan dan diskrit. Namun demikian untuk penelitian ini yang digunakan hanya atribut diskrit sebab apabila memakai atribut berkelanjutan, hasil keluaran dari pohon keputusan dapat panjang dan tidak efisien[4].
- Hasil pohon keputusan C5.0 bisa dipotong atau terdapat pruning. Atribut dengan nilai *information gain* teratas akan terpilih menjadi parent untuk node berikutnya. Algoritma ini membuat pohon keputusan melalui cara pemisahan dan mendominasi sampel secara rekursif[4].

Ini merupakan rumus untuk mencari *informasi gain*

$$I\ (s1\ ,\ s2\ ,\ \dots\ ,\ sm) = -\sum_{i=1}^{m} \text{pi log2 (pi)}$$

Yang akan menjadi *operand* pengurang data *entropy* untuk mencari *Gain*, gain yang tertinggi akan membaentuk pohon dari akar, daun dan seterusnya

Ini rumus untuk mencari nilai *entropy*

$$E\ (A) = \sum_{j=1}^{y} \frac{s1j+\dots+smj}{s}\ I\ (S1j, \dots, Smj)$$

$$\frac{s1j+\dots+smj}{s}$$

Berikut rumus untuk mencari nilai *Gain*

$$Gain\ (A) = I\ (s_1\ ,\ s_2\ ,\ \dots\ ,\ s_m) - E(A)$$

E. Nearest-Neighbor Classifer

K-Nearest Neighbor Algorithm sama halsnya decision tree, K-Nearest Neighbor sangat sering dipakai untuk klasifikasi melalui target dari algoritme ini ialah untuk mengklasifikasi objek baru melalui atribut dan training samples. Algoritme k-nearest neighbor (k-NN atau KNN) yaitu sebuah metode untuk mengklasifikasi objek melalui data pembelajaran yang jaraknya sangat dekat dengan objek tersebut. Dalam hal tersebut, jumlah tetangga terdekat ditetapkan dari *user* yang dinyatakan dengan *k*. sebagai contoh nilai *k*=5, maka setiap *data testing* dihitung jaraknya terhadap *data training* dan dipilih 5 *data training* yang jaraknya sangat dekat ke *data testing*[5].

Untuk mendapatkan jarak antara dua titik yaitu titik pada *data training* (x) dan titik pada *data testing* (y) maka dilakukan rumus Euclidean[5].

$$d_{(x,y)}\sqrt{\sum_{i=1}^{n}(xi - yi)^2}$$

d adalah jarak antara titik pada data *training x* dan titik data *testing y* yang akan diklasifikasi, yang mana $x=x_1,x_2,\dots,x_i$ dan $y=y_1,y_2,\dots,y_i$ dan *I* merepresentasikan nilai atribut sedangkan *n* merupakan dimensi atribut[5].

III. PERANCANGAN SISTEM

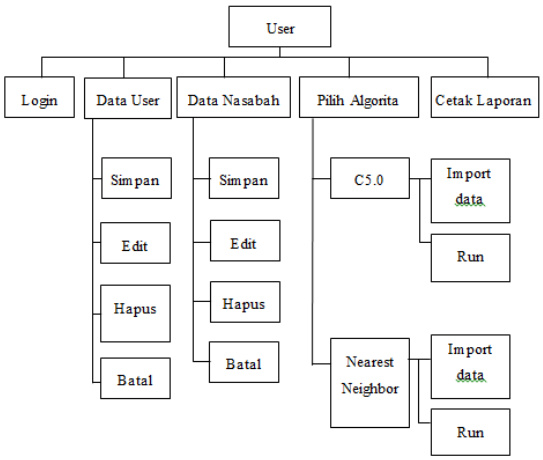
Perancangan sistem meliputi Perancangan Arsitektur, Perancangan Prosedural dan Perancangan Antarmuka.

A. Perancangan Arsitektur

Dengan perancangan arsitekur ini mampu memudahkan dalam pembanguan aplikasi yang akan dibangun

- Perancangan Struktur

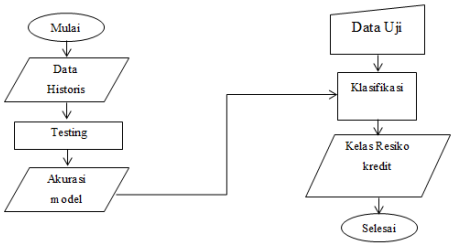
Dalam perancangan ini berisikan menu dan submenu yang berfungsi untuk memudahkan pengguna dalam menggunakan aplikasi tersebut. Berikut ini digambarkan mengenai struktur menu dalam aplikasi ini:



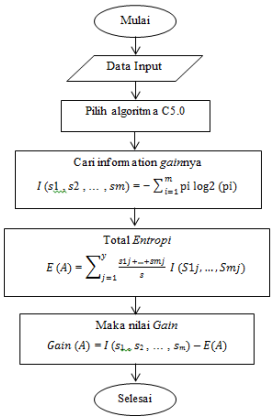
Gambar 1 Perancangan Struktur Menu

B. Pengujian Prosedural

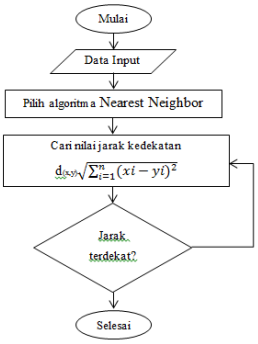
Perancangan ini akan membahas mengenai detail algoritma yang akan dinyatakan kedalam suatu program.



Gambar 2 Flowchar sistem secara keseluruhan



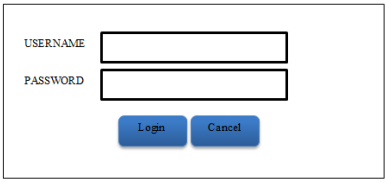
Gambar 3 Flowchar tahap klasifikasi dengan algoritma C5.0



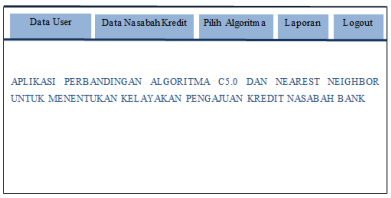
Gambar 4 Flowchar tahap klasifikasi dengan algoritma Nearest Neighbor

C. Perancaangan Antarmuka

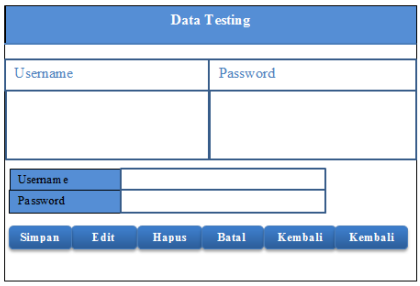
Perancangan adalah suatu bentuk kasar dari program yang akan dibuat untuk kebutuhan *interface* dengan pengguna.



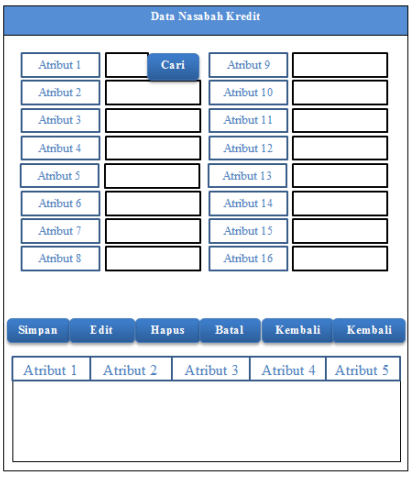
Gambar 5 Perancangan form login



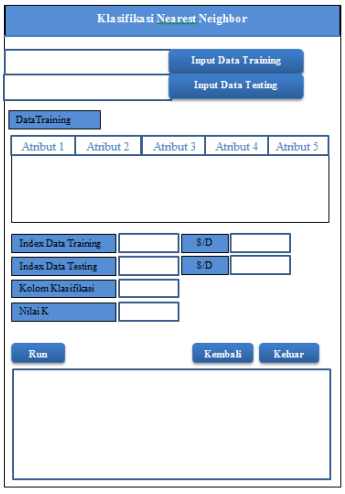
Gambar 6 Perancangan form menu utama



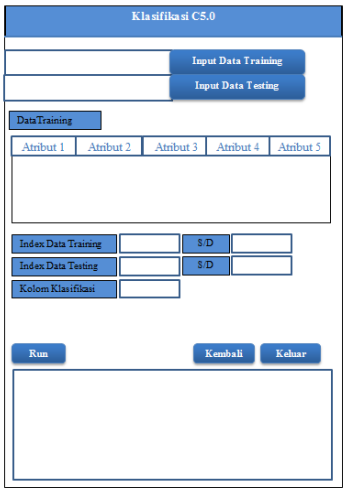
Gambar 7 Perancangan form Data User



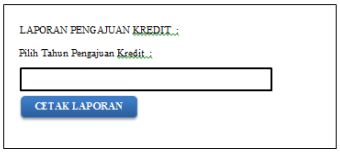
Gambar 8 Perancangan form data nasabah kredit



Gambar 9 Perancangan form klasifikasi Neaest Neighbor



Gambar 10 Perancangan form klasifikasi C5.0



Gambar 11 Perancangan form untuk cetak laporan

IV. PENGUJIAN

Pada bagian ini akan dilakukan pengujian pada sistem yang sudah dirancang.

.

A. Pengujian Antarmuka

Pada tampilan ini dilakukan penerapan hasil perancangan antarmuka ke dalam sistem yang dibangun

1. Tampilan Menu Login

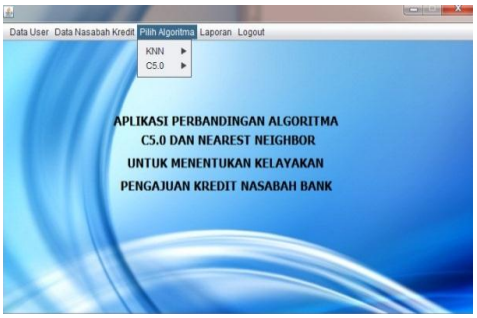
Ini merupakan tampilan awal aplikasi, Pada tampilan ini user akan mengisi username dan password untuk mengecek validasi Identitas, dan ketika berhasil maka aplikasi ini dapat diakses.



Gambar 12 Tampilan Menu Login

2. Tampilan Menu Utama

Tampilan ini akan menampilkan menu-menu yang merupakan fungsi-fungsi dari aplikasi



Gambar 13 Tampilan Menu Utama

3. Tampilan Data Nasabah Kredit

Tampilan ini merupakan tampilan antarmuka untuk menampilkan data nasabah kredit.



Gambar 14 Form Data Nasabah Kredit

4. Tampilan Data User

Form ini merupakan antarmuka untuk menambahkan data user yang nantinya dapat digunakan untuk login.



Gambar 15Form Data User

5. Tampilan Proses C5.0 dan Nearest Neighbor

Form ini merupakan proses dari perhitungan C5.0 maupun Nearest Neighbor, dimana nasabah tersebut layak atau tidak layak untuk menerima kredit.

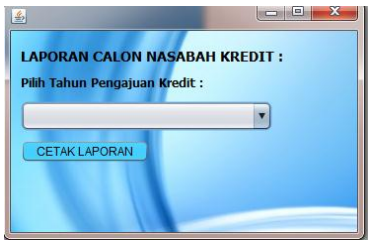


Gambar 16 Form Proses Klasifikasi Algoritma Nearest Neighbor



Gambar 17 Form Proses Klasifikasi Algoritma C5.0

Form ini merupakan antarmuka untuk mencetak hasil laporan dari data nasabah yang telah di proses persetujuan pengajuan kredit.



Gambar 18 Form Cetak Laporan

B. Pengujian Sistem

Pengujian ini merupakan pengujian yang dilakukan pada aplikasi yang dibuat, apakah aplikasi ini berhasil atau tidak.

Untuk mengetahui tingkat keberhasilan pada sistem ini maka digunakan rumus di bawah ini :

$$tingkat\ keberhasilan = \frac{\sum hasil\ pengujian\ berjumlah\ benar(kecocokan)}{\sum banyaknya\ data\ sampel} x100\%$$

Rumus ini memperhitungkan kesamaan data antara data sesungguhnya dengan data hasil prediksi yang dihasilkan oleh aplikasi.

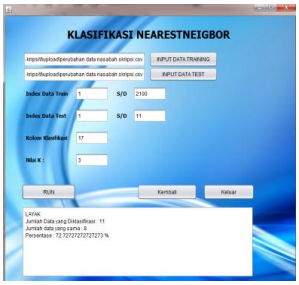
1. Pengujian 1 Database (Testing Data = 11 Data)



Gambar 19 Klasifikasi Nearest Neighbor Testing 11 Data (k = 1)

Dapat terlihat kecocokan antara hasil prediksi yang telah di proses dengan menggunakan algoritma Nearest Neighbor K = 1. Pada data hasil testing terdapat 11 data yang cocok dari 11 data yang diuji dan 0 data yang tidak cocok, untuk presentase keberhasilannya adalah :

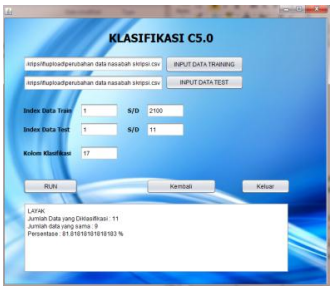
Tingkat keberhasilan (11/11)*100% = 100 %



Gambar 20 Klasifikasi Nearest Neighbor Testing 11 Data (k = 3)

Dapat terlihat kecocokan antara hasil prediksi yang telah di proses dengan menggunakan algoritma Nearest Neighbor K = 3. Pada data hasil testing terdapat 8 data yang cocok dari 11 data yang diuji dan 3 data yang tidak cocok, untuk presentase keberhasilannya adalah :

Tingkat keberhasilan (8/11)*100% = 72,72727273 %

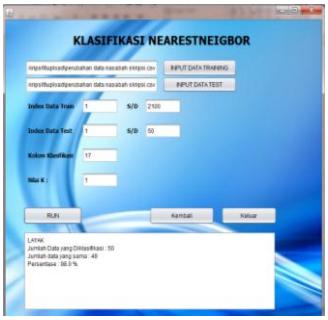


Gambar 21 Klasifikasi C5.0 Testing 11 Data

Dapat terlihat kecocokan antara hasil prediksi yang telah di proses dengan menggunakan algoritma C5.0. Pada data hasil testing terdapat 8 data yang cocok dari 11 data yang diuji dan 3 data yang tidak cocok, untuk presentase keberhasilannya adalah :

Tingkat keberhasilan (8/11)*100% = 72,72727273 %

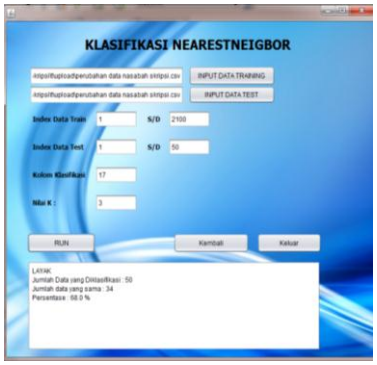
2. Pengujian 1 Database (Testing Data = 50 Data)



Gambar 22 Klasifikasi Nearest Neighbor Testing 50 Data (k = 1)

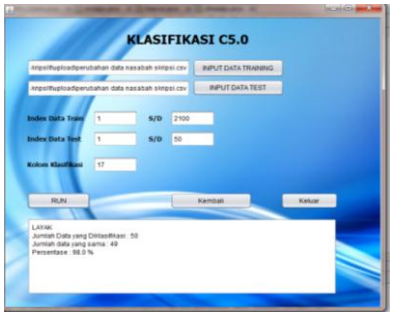
Dapat terlihat kecocokan antara hasil prediksi yang telah di proses dengan menggunakan algoritma Nearest Neighbor K = 1. Pada data hasil testing terdapat 49 data yang cocok dari 50 data yang diuji dan 1 data yang tidak cocok, untuk presentase keberhasilannya adalah :

Tingkat keberhasilan (49/50)*100% = 98,0 %



Gambar 23 Klasifikasi Nearest Neighbor Testing 50 Data (k = 3)

dapat terlihat kecocokan antara hasil prediksi yang telah di proses dengan menggunakan algoritma Nearest Neighbor K = 3. Pada data hasil testing terdapat 34 data yang cocok dari 50 data yang diuji dan 16 data yang tidak cocok, untuk presentase keberhasilannya adalah :
Tingkat keberhasilan (34/50)*100% = 68,0 %



Gambar 24 Klasifikasi C5.0 Testing 11 Data

dapat terlihat kecocokan antara hasil prediksi yang telah di proses dengan menggunakan algoritma C5.0. Pada data hasil testing terdapat 49 data yang cocok dari 50 data yang diuji dan 1 data yang tidak cocok, untuk presentase keberhasilannya adalah :

Tingkat keberhasilan (49/50)*100% = 98,0 %

3. Tabel Hasil Pengujian

Pada bagian ini pengujian data perbandingan ditampilkan secara keseluruhan pada tabel.

Tabel 1 Hasil Penelitian

Jumlah Data Training	Jumlah Data Testing	Persentase C5.0	Persentase Nearest Neighbor	
			K = 1	K = 3
2100	11	72,7273 %	100 %	72,7273 %
2100	50	98,0 %	98,0 %	68,0 %
2100	100	99,0 %	99,0 %	76,0 %
2100	501	85,02994%	99,800399 %	85,02994%

V. Kesimpulan dan Saran

Kesimpulan

Setelah melakukan analisis, peranvcangan, dan pengujian, maka dapat diperoleh kesimpulan sebagai berikut :

Setelah melakkan analisis, peranvcangan, dan pengujian, maka dapat diperoleh kesimpulan sebagai berikut :

1. Metode klasisfikasi dengan menggunakan kedua algoritma yaitu C5.0 dan Nearest Neighbor dapat digunakan untuk melakukan prediksi kelayakan pengajuan kredit dengan cepat.
2. Jika data yang dimiliki atau data training nol, Kedua algoritma yaitu C5.0 dan Nearest Neighbor tidak dapat mengklasifikasikan record, sehingga data yang di prediksi akan bernilai nol atau tidak layak.
3. Dari dua algoritma yang diuji pada bab 4, terlihat bahwa pada data kecil C5.0 mampu seimbang dengan Nearest Neighbor dengan k = 1, dan lebih baik dari Nearest Neighbor dengan k = 3. Akan tetapi ketika data berjumlah besar yaitu lebih dari 100 data, maka persentase C5.0 hanya dapat seimbang dan bahkan lebih rendah dari Nearest Neighbor k = 3, dan ketika dibandingkan dengan Nearest Neighbor k = 1, C5.0 memiliki persentase lebih rendah.

Saran

Jika data historis yang di dalam data base berjumlah banyak, maka dapat dibuat kemungkinan pertimbangan baru yaitu menambah kelas output. Data harus lebih bervariasi untuk menghindari adanya data traning nol, yang akan menyebabkan data yang di uji menjadi bernilai nol.

DAFTAR PUSTAKA

[1] OJK. (2013). *Undang-undang Nomor 7 Tahun 1992 tentang Perbankan Sebagaimana Diubah dengan Undang-undang Nomor 10 Tahun 1998*. Diakses: 9 april 2017. Site: http://www.ojk.go.id/id/kanal/perbankan/regulasi/undang-undang/Pages/undang-undang-nomor-7-tahun-1992-tentang-perbankan-sebagaimana-diubah-dengan-undang-undang-nomor-10-tahun-1998.aspx

[2] Hernawan, K.A. (2015). *Aplikasi Data Mining Menggunakan Naïve Bayes Classifier Untuk Persetujuan Pengajuan Kredit.* Fakultas Teknik Dan Ilmu Komputer. Bandung : Universitas Komputer Indonesia

[3] Hermawati, F. A. (2013). *Data Mining*. Yogyakarta : ANDI.

[4] Ernawati, I. (2008). *Prediksi Status Keaktifan Studi Mahasiswa Dengan Algoritma C5.0 dan K-Nearest Neighbor*. Sekolah Pasca Sarjana. Bogor: Institut Pertanian Bogor

[5] Kusrini, Hartati. S., Wardoyo. R., Harjoko. A. (2009), *Perbandingan metode nearest neighbor dan Algoritma c4.5 untuk menganalisis Kemungkinan pengunduran diri calon Mahasiswa di stmik amikom Yogyakarta.* Yokyakarta : STMIK AMIKOM