
**KLASIFIKASI TINGKAT KELANCARAN NASABAH DALAM
MEMBAYAR PREMI DENGAN MENGGUNAKAN METODE
K-NEAREST NEIGHBOR DAN ANALISIS DISKRIMINAN FISHER
(Studi kasus: Data Nasabah PT. Prudential Life Samarinda
Tahun 2019)**

Amanah^{1,3}, Memi Nor Hayati², Rito Goejantoro³

¹ Mahasiswa Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Mulawarman

^{2,3} Dosen Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas
Mulawarman

Email : amanahadam44@gmail.com

ABSTRACT

Classification is a technique to form a model of data that is already known to its classification group. The model that was formed will be used to classify new objects. The K-Nearest Neighbor (K-NN) algorithm is a method for classifying new objects based on their K nearest neighbor. Fisher discriminant analysis is a multivariate technique for separating objects in different groups to form a discriminant function for allocate new objects in groups. This research has a goal to determine the results of classifying customer premium payment status using the K-NN method and Fisher discriminant analysis and comparing the accuracy of the K-NN method classification and Fisher discriminant analysis on the insurance customer premium payment status. The data used is the insurance customer data of PT. Prudential Life Samarinda in 2019 with current premium payment status or non-current premium payment status and four independent variables are age, duration of premium payment, income and premium payment amount. The results of the comparative measurement of accuracy from the two analyzes show that the K-NN method has a higher level of accuracy than Fisher discriminant analysis for the classification of insurance customers premium payment status. The results of misclassification using the APER (Apparent Error Rate) in K-NN method is 15% while in Fisher discriminant analysis is 30%.

Keywords: Fisher discriminant analysis, Insurance, K-Nearest Neighbor.

PENDAHULUAN

Dalam kehidupan sehari-hari, manusia selalu memiliki risiko dalam setiap kegiatan yang dilakukan. Untuk meminimalisir tingkat risiko yang ada, perlu adanya jaminan atau asuransi. Salah satu jenis asuransi yang paling diminati oleh masyarakat adalah asuransi jiwa. Asuransi jiwa adalah usaha kerjasama dari sejumlah orang yang sepakat memikul kesulitan keuangan bila terjadi musibah terhadap salah satu anggotanya.

Dalam menentukan premi nasabah dalam suatu polis asuransi, perusahaan

asuransi tentunya tidak ingin mengalami kerugian. Perusahaan asuransi jiwa memiliki kriteria-kriteria tertentu di mana nasabah tersebut dapat memiliki polis asuransi jiwa. Kriteria calon pemilik polis asuransi jiwa dapat dilihat dari umur, pendapatan, status pekerjaan, masa pembayaran, besar pembayaran premi, dan lain-lain [1]. Pengelompokkan nasabah dalam kategori pembayaran premi dengan kriteria tertentu dapat dilakukan menggunakan analisis statistika yaitu dengan metode klasifikasi.

Klasifikasi merupakan proses penemuan model atau fungsi yang menjelaskan atau membedakan konsep pada kelompok data dengan tujuan untuk dapat memperkirakan kelompok dari suatu objek yang labelnya tidak diketahui [2]. Metode klasifikasi yang umum digunakan antara lain *Decision Tree*, *K-Nearest Neighbor*, *Naïve Bayes*, *Neural Network*, dan lain sebagainya. Selain metode tersebut, terdapat metode analisis lain di mana pada proses pengklasifikasiannya memperhatikan asumsi tertentu yaitu analisis diskriminan [3].

Algoritma *K-Nearest Neighbor* (K-NN) dilakukan dengan mencari kelompok K objek dalam data *training* yang paling dekat (mirip) dengan objek pada data baru atau data *testing* [4]. Analisis diskriminan merupakan suatu metode statistika yang digunakan untuk mengklasifikasikan suatu individu atau objek ke dalam suatu kelompok yang telah ditentukan sebelumnya berdasarkan variabel-variabel bebasnya [5].

Sebuah perusahaan asuransi pastilah mempunyai data yang begitu besar. Banyak yang belum menyadari bahwa dari pengolahan data-data tersebut dapat memberikan informasi berupa klasifikasi data nasabah yang akan bergabung pada perusahaan itu sendiri. Berdasarkan latar belakang tersebut, penulis tertarik untuk melakukan penelitian ilmiah dengan judul “Klasifikasi Tingkat Kelancaran Nasabah dalam Membayar Premi dengan Menggunakan Metode *K-Nearest Neighbor* dan Analisis Diskriminan Fisher (Studi Kasus: Data Nasabah PT. *Prudential Life Samarinda Tahun 2019*)”.

METODOLOGI PENELITIAN

Sumber Data dan Variabel Penelitian

Penelitian ini menggunakan data sekunder yang diperoleh dari PT. Prudential Life Jalan MT Haryono Samarinda. Variabel yang digunakan dalam penelitian ini adalah sebagai berikut:

1. Status Pembayaran Premi (Y)

Status pembayaran premi nasabah dengan kategori status pembayaran premi lancar dinotasikan sebagai Y_1 dan kategori status pembayaran premi tidak lancar dinotasikan sebagai Y_2 .

2. Usia (X_1)

Usia bertanggung pada saat proses pengambilan data dalam satuan tahun.

3. Lama Pembayaran Premi (X_2)

Lama pembayaran premi bagi bertanggung yang dihitung sejak awal mendaftar di PT. Prudential Life Samarinda hingga saat pengambilan data dalam satuan tahun.

4. Pendapatan (X_3)

Pendapatan bertanggung setiap bulannya pada saat mendaftar di PT. Prudential Life Samarinda dalam satuan Rupiah.

5. Besar Pembayaran Premi (X_6)

Besar pembayaran premi bagi bertanggung yang dibayarkan setiap bulannya dalam satuan Rupiah.

Pengolahan data pada penelitian ini menggunakan metode K-NN dan analisis diskriminan Fisher.

Metode Analisis

Tahapan dari klasifikasi menggunakan metode K-NN dalam penelitian ini adalah sebagai berikut:

a. Standarisasi Data

$$\hat{x}_{(j,g)} = \frac{x_{(j,g)} - \bar{x}_g}{s_g} \quad (1)$$

b. Penentuan Nilai K

Dalam penelitian ini, nilai K yang digunakan adalah 1, 3, 5, 7 dan 9.

c. Perhitungan jarak antar data *training* dan data *testing* menggunakan jarak Euclid

$$d_{(a,b)} = \sqrt{\sum_{g=1}^p (x_{ag} - y_{bg})^2} \quad (2)$$

d. Pengurutan jarak hasil perhitungan

e. Proses klasifikasi berdasarkan nilai K

f. Perhitungan akurasi menggunakan nilai APER

$$APER = \frac{\text{jumlah obyek salah klasifikasi}}{\text{jumlah prediksi yang dilakukan}} \times 100\% \quad (3)$$

Adapun langkah-langkah yang dilakukan dalam mengolah data pengklasifikasian dengan analisis diskriminan Fisher adalah sebagai berikut:

- a. Melakukan pengujian asumsi normalitas multivariat menggunakan persamaan (4) dan kesamaan matriks kovarians dengan uji *Box's M*.

$$(\mathbf{x}_j - \bar{\mathbf{x}})^T \mathbf{S}^{-1} (\mathbf{x}_j - \bar{\mathbf{x}}) \leq \chi_p^2(0,5) \quad (4)$$

- b. Menentukan fungsi diskriminan Fisher dua kelompok dengan menggunakan data *training*.

$$y = \mathbf{b}^T \mathbf{X} \quad (5)$$

dengan,

$$\mathbf{b} = \mathbf{S}^{-1} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) \quad (6)$$

- c. Menghitung nilai batas kelompok (\hat{m})

$$\hat{m} = \frac{1}{2} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)^T \mathbf{S}^{-1} (\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_2) \quad (7)$$

- d. Melakukan perhitungan skor diskriminan menggunakan data *testing*.

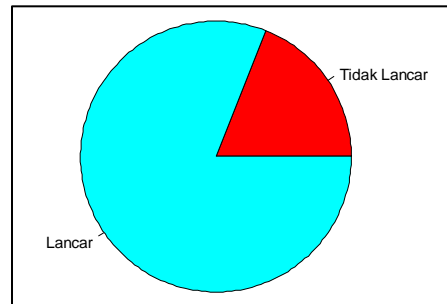
- e. Pengukuran tingkat akurasi menggunakan nilai APER berdasarkan persamaan (3).

HASIL DAN PEMBAHASAN

Sebelum memasuki proses klasifikasi, dilakukan pengacakan data terlebih dahulu. Setelah itu data dibagi dengan proporsi data *training* dan *testing* 70:30 untuk kemudian dilakukan proses klasifikasi. Adapun variabel terikat dalam penelitian ini adalah status pembayaran premi nasabah asuransi dengan kategori lancar dan tidak lancar dengan variabel bebas adalah pendapatan, usia, lama pembayaran premi dan besar pembayaran premi.

Statistika Deskriptif

Berikut ini merupakan gambaran dari variabel terikat yaitu status pembayaran premi.



Gambar 1. Persentase Status Pembayaran Premi

Dari 95 data nasabah asuransi dapat diketahui bahwa 81,05% atau sebanyak 77 orang nasabah asuransi memiliki status pembayaran premi lancar dan 18,95% atau sebanyak 18 orang nasabah asuransi memiliki status pembayaran premi tidak lancar.

Tabel 1. Statistika Deskriptif Nasabah Kategori Lancar

Statistika Deskriptif	Variabel			
	X_1	X_2	X_3	X_4
Rata-rata	39,9	7,1	9.298.701	698.701,3
Simpangan Baku	8,1	1,7	5.598.555	367.600,1
Minimum	24	2	4.000.000	300.000
Maksimum	60	10	25.000.000	2.000.000

Tabel di atas merupakan gambaran umum variabel usia, lama pembayaran premi, pendapatan dan besar pembayaran premi nasabah dengan status pembayaran premi kategori lancar.

Tabel 2. Statistika Deskriptif Nasabah Kategori Tidak Lancar

Statistika Deskriptif	Variabel			
	X_1	X_2	X_3	X_4
Rata-rata	41,6	6,7	6.944.444	575.000
Simpangan Baku	8,9	1,8	4.249.183	198.709,1
Minimum	26	4	600.000	350.000
Maksimum	58	9	20.000.000	1.000.000

Tabel di atas merupakan gambaran umum variabel usia, lama pembayaran premi, pendapatan dan besar pembayaran premi nasabah dengan status pembayaran premi kategori tidak lancar.

Klasifikasi Status Pembayaran Premi Menggunakan Metode K-NN

Langkah pertama yang dilakukan adalah standarisasi data. Hal ini dilakukan agar semua variabel berada dalam jangkauan yang sama sehingga proporsi pengaruh pada fungsi di dalam klasifikator dapat seimbang. Standarisasi data dihitung menggunakan persamaan (1).

Selanjutnya menghitung Jarak Euclid yang digunakan untuk mencari jarak antara data *training* dengan data *testing* menggunakan persamaan (2). Berikut adalah perhitungan jarak Euclid untuk data *testing* pertama (sampel A-119). Perhitungan jarak Euclid dilakukan hingga data *training* ke-95 dengan data *testing* ke-40:

$$\begin{aligned}
 d_{(1,1)} &= \sqrt{(x_{11} - y_{11})^2 + \dots + (x_{14} - y_{14})^2} \\
 &= \sqrt{((-1,1) - (-1,6))^2 + \dots + ((-0,9) - (-0,5))^2} \\
 &= 2,4
 \end{aligned}$$

Setelah dilakukan perhitungan jarak Euclid antara data *training* dengan data *testing*, langkah selanjutnya yaitu mengurutkan jarak Euclid kemudian dilanjutkan dengan mencari K tetangga terdekat menggunakan K awal yaitu 1, 3, 5, 7 dan 9. Pada penggunaan 1-NN, prediksi klasifikasinya adalah label kelas 1 (status pembayaran premi lancar) karena label kelas 1 adalah klasifikasi yang paling banyak muncul dari *rank* pertama. Pada penggunaan 3-NN prediksi klasifikasinya adalah label kelas 1 (status pembayaran premi lancar) karena label kelas 1 adalah klasifikasi yang paling

Data Testing	Prediksi K-NN					Klasifikasi Pada Data Asli
	1	3	5	7	9	
A-119	1	1	1	1	1	1
A-129	2	1*	1*	1*	1*	2
A-132	1	1	1	1	1	1
⋮	⋮	⋮	⋮	⋮	⋮	⋮
A-061	1	1	1	1	1	1
A-100	1	1	1	1	1	1
Prediksi Salah	11	10	9	7	6	

banyak muncul dari *rank* pertama sampai dengan *rank* ketiga, dan seterusnya hingga prediksi dengan menggunakan K = 9. Selanjutnya dilakukan hal yang sama untuk

data *testing* kedua, ketiga dan seterusnya hingga data *testing* ke-40 (Sampel A-100) sehingga mendapatkan hasil prediksi. Ringkasan hasil prediksi klasifikasi menggunakan K-NN dapat dilihat pada tabel di bawah ini.

Tabel 3. Perbandingan Hasil Prediksi

Berdasarkan tabel di atas, dapat dilihat bahwa sel dengan angka yang diberi tanda (*) memiliki perbedaan prediksi klasifikasi dengan data aslinya.

Pengukuran tingkat akurasi metode K-NN dilakukan dengan menghitung peluang kesalahan klasifikasi. Semakin kecil kesalahan klasifikasi menunjukkan bahwa semakin baik hasil klasifikasi yang didapatkan. Berdasarkan persamaan (3) maka diperoleh nilai APER untuk K-NN dengan K = 1 sebagai berikut:

$$\begin{aligned}
 \text{APER} &= \frac{6+5}{40} \times 100\% \\
 &= \frac{11}{40} \times 100\% \\
 &= 27,5\%
 \end{aligned}$$

Dengan cara yang sama, peneliti juga menghitung nilai APER dari hasil prediksi klasifikasi menggunakan nilai K yang lain yaitu 3, 5, 7 dan 9. Nilai APER untuk setiap K pada metode K-NN dapat dilihat pada Tabel 4.

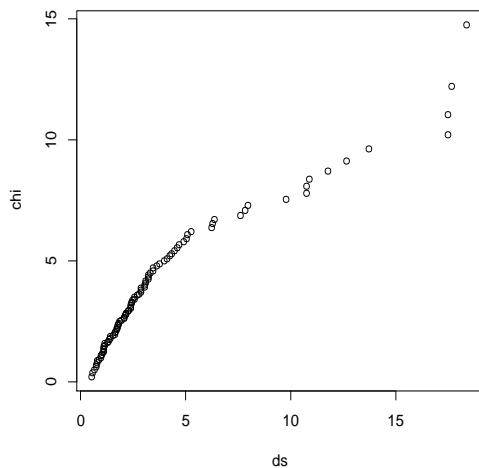
Tabel 4. Perbandingan Nilai APER pada K-NN

K-NN	APER
1-NN	27,5%
3-NN	25%
5-NN	22,5%
7-NN	17,5%
9-NN	15%

Berdasarkan tabel di atas dapat diketahui nilai APER yang menunjukkan kesalahan klasifikasi dalam memprediksi kelas data *testing*. Nilai APER yang paling kecil berada pada 9-NN, sehingga nilai APER 9-NN yang akan dibandingkan dengan nilai APER pada hasil pengklasifikasian menggunakan analisis diskriminan Fisher.

Klasifikasi Status Pembayaran Premi Menggunakan Analisis Diskriminan Fisher

Sebelum dilakukan klasifikasi menggunakan analisis diskriminan Fisher, data harus memenuhi asumsi normal multivariat dan matriks kovarians antar kelompok homogen. Menurut persamaan (4) pengecekan normal multivariat dapat menggunakan perhitungan jarak kuadrat (d_j^2), di mana minimal setengah dari nilai d_j^2 kurang dari atau sama dengan $\chi_{(4;0,5)}^2 = 3,357$. Dari 95 nilai d_j^2 terdapat 63 nilai yang kurang dari 3,357. Hal ini menandakan bahwa proporsi nilai d_j^2 yang kurang dari 3,357 adalah sebesar 0,66 atau 66%. Sehingga, dapat disimpulkan bahwa data keempat variabel bebas berdistribusi normal multivariat.



Gambar 2. Q-Q plot uji normalitas multivariat

Pada gambar di atas terlihat bahwa plot cenderung membentuk garis lurus, sehingga dapat disimpulkan bahwa data keempat variabel bebas (pendapatan, usia, lama pembayaran premi dan besar pembayaran premi) berdistribusi normal multivariat.

Selanjutnya, pengujian kesamaan matriks kovarians antar kelompok dilakukan dengan menggunakan uji Box's M. Hasil pengujian menggunakan uji Box's M dapat dilihat pada tabel di bawah ini.

Tabel 5. Uji Box's M

C	$\chi_{0,05;10}^2$	Keputusan
---	--------------------	-----------

13,03	18,31	gagal tolak H_0
-------	-------	-------------------

Berdasarkan tabel di atas diperoleh keputusan bahwa H_0 gagal ditolak yang artinya matriks kovarians antar kelompok data nasabah asuransi adalah sama.

Klasifikasi status pembayaran premi data nasabah asuransi menggunakan analisis diskriminan Fisher dilakukan dengan mengetahui fungsi diskriminan Fisher menggunakan data *training*. Langkah pertama dilakukan perhitungan matriks rata-rata kelompok dan didapat matriks rata-rata kelompok 1 dan kelompok 2 sebagai berikut.

$$\bar{\mathbf{x}}_1 = \begin{bmatrix} 39,91 \\ 7,17 \\ 9.298.701 \\ 698.701,3 \end{bmatrix} \quad \bar{\mathbf{x}}_2 = \begin{bmatrix} 41,61 \\ 6,67 \\ 6.944.444 \\ 575.000 \end{bmatrix}$$

Selanjutnya menghitung selisih dan jumlah matriks rata-rata kelompok sehingga didapat hasil seperti di bawah ini.

$$\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2 = \begin{bmatrix} -1,70 \\ 0,50 \\ 2.354.257 \\ 123.701,3 \end{bmatrix} \quad \bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_2 = \begin{bmatrix} 81,52 \\ 13,84 \\ 16.243.150 \\ 1.273.701 \end{bmatrix}$$

Nilai matriks kovarians keempat variabel bebas pada setiap kelompok dihitung sehingga diperoleh matriks kovarians kelompok 1 dan kelompok 2. Selanjutnya, menentukan matriks kovarians gabungan.

$$\mathbf{S} = \left(\frac{n_1 - 1}{(n_1 - 1) + (n_2 - 1)} \right) \mathbf{S}_1 + \left(\frac{n_2 - 1}{(n_1 - 1) + (n_2 - 1)} \right) \mathbf{S}_2$$

$$= \begin{bmatrix} 67,56 & 4,83 & -2.832.957 & -18.897,7 \\ 4,83 & 2,92 & -1.374.643 & -30.139,5 \\ -2.832.957 & -1.374.643 & 2,86 \times 10^{13} & 1,12 \times 10^{12} \\ -18.897,7 & -30.139,5 & 1,12 \times 10^{12} & 1,16 \times 10^{11} \end{bmatrix}$$

Selanjutnya, menghitung nilai koefisien dari masing-masing variabel menggunakan persamaan (6) sehingga didapat fungsi diskriminan Fisher sebagai berikut:

$$y = -0,042X_1 + 0,28X_2 + 7,62 \times 10^{-8} X_3 + 3,94 \times 10^{-7} X_4 \quad (8)$$

Nilai batasan kelompok (\hat{m}) dapat dihitung menggunakan persamaan (7) dan didapat $\hat{m} = 1,103$.

Tahap terakhir adalah dilakukan perhitungan nilai fungsi diskriminan Fisher (y) pada masing-masing data *testing*. Pada data *testing* pertama diketahui nasabah asuransi memiliki usia (X_1) 27 tahun, lama pembayaran premi (X_2) 5 tahun, pendapatan (X_3) sebesar Rp. 5.000.000,00 dan besar pembayaran premi (X_4) sebesar Rp. 500.000,00. Keempat nilai variabel bebas tersebut kemudian di substitusikan terhadap fungsi diskriminan Fisher. Nilai fungsi diskriminan pada data *testing* pertama menggunakan persamaan (8) sebagai berikut:

$$y = -0,042X_1 + 0,28X_2 + 7,62 \times 10^{-8} X_3 + 3,94 \times 10^{-7} X_4$$

$$\hat{y}_1 = -0,042(27) + 0,28(5) + 7,62 \times 10^{-8}(5.000.000) + 3,94 \times 10^{-7}(500.000)$$

$$\hat{y}_1 = 0,271$$

Sehingga dapat diketahui nilai fungsi diskriminan Fisher pada data *testing* pertama sebesar 0,271. Berdasarkan aturan analisis diskriminan Fisher dapat disimpulkan nilai $\hat{y}_1 = 0,271 < \hat{m} = 1,103$ maka obyek pengamatan data *testing* pertama dapat diklasifikasikan ke dalam kelompok kedua yaitu nasabah asuransi memiliki status pembayaran premi tidak lancar. Cara yang sama dilakukan untuk semua data *testing* yang ada untuk mengetahui hasil klasifikasinya.

Pengukuran tingkat akurasi analisis diskriminan Fisher dilakukan dengan menghitung peluang kesalahan klasifikasi menggunakan APER.

$$\begin{aligned} \text{APER} &= \frac{8+4}{40} \times 100\% \\ &= \frac{12}{40} \times 100\% \\ &= 30\% \end{aligned}$$

Tingkat akurasi pada metode KNN dibandingkan dengan tingkat akurasi pada analisis diskriminan Fisher.

Tabel 6. Perbandingan Tingkat Akurasi

Metode	APER
Metode K-NN (dengan K optimal = 9)	15%
Analisis Diskriminan Fisher	30%

Berdasarkan tabel di atas dapat diketahui bahwa pada metode K-NN (dengan K optimal = 9) menunjukkan kesalahan klasifikasi dalam memprediksi kelas data *testing* sebesar 15% dan pada analisis diskriminan Fisher menunjukkan kesalahan klasifikasi dalam memprediksi kelas data *testing* sebesar 30%. Hal ini menunjukkan bahwa tingkat akurasi pengklasifikasian metode K-NN (dengan K optimal = 9) lebih baik dibandingkan dengan analisis diskriminan Fisher dalam mengklasifikasikan data nasabah PT Prudential Life Samarinda Tahun 2019.

KESIMPULAN

Berdasarkan hasil analisis data dan pembahasan diperoleh kesimpulan sebagai berikut:

1. Pengklasifikasian status pembayaran premi pada data nasabah asuransi PT. Prudential Life Samarinda Tahun 2019 menggunakan metode K-NN (dengan K optimal = 9) diperoleh hasil yaitu 34 nasabah tepat diklasifikasikan dan sisanya 6 nasabah tidak tepat diklasifikasikan.
2. Pengklasifikasian status pembayaran premi pada data nasabah asuransi PT. Prudential Life Samarinda Tahun 2019 menggunakan analisis diskriminan Fisher diperoleh hasil yaitu 26 nasabah tepat diklasifikasikan dan sisanya 14 nasabah tidak tepat diklasifikasikan.

3. Hasil pengukuran tingkat akurasi klasifikasi status pembayaran premi pada data nasabah asuransi PT. Prudential Life Samarinda Tahun 2019 menggunakan APER diperoleh hasil kesalahan klasifikasi metode K-NN (dengan K optimal = 9) adalah sebesar 15% dan analisis diskriminan Fisher sebesar 30%. Hal ini menunjukkan bahwa metode K-NN (dengan K optimal = 9) memberikan ketepatan prediksi klasifikasi yang lebih baik dalam mengklasifikasikan data nasabah PT Prudential Life Samarinda Tahun 2019.

DAFTAR PUSTAKA

- [1] Prakoso, D. (1994). *Asuransi Indonesia*. Semarang: Dahara Prize.
- [2] Prasetyo, E. (2014). *Data Mining: Konsep dan Aplikasi Menggunakan Matlab*. Yogyakarta: Andi Offset.
- [3] Artha, C. U, Nasution, Y.N dan Purnamasari, I. (2016). Perbandingan Hasil Klasifikasi Menggunakan Regresi Logistik dan Analisis Diskriminan Kuadratik Pada Kasus Pengklasifikasian Jurusan Di SMA Negeri 8 Samarinda Tahun Ajaran 2014/2015. *Jurnal Eksponensial*. 7(2). 179-186.
- [4] Jayanti, R. D dan Noeryanti. (2014). Aplikasi Metode K-NN dan Analisis Diskriminan untuk Analisis Resiko Kredit pada Koperasi Simpan Pinjam di Kopinkra Sumber Rejeki. *Jurnal SNAST*. Vol 4 No.8 275-284.
- [5] Dillon, W. R and Goldstein, M. (1984). *Multivariate Analysis Methods and Application*. New York: John Wiley & Sons. Inc.