

Penentuan Status Tahapan Keluarga Sejahtera Kecamatan Sidareja Menggunakan Teknik Data Mining

R Bagus Bambang Sumantri¹, Ema Utami²

STMIK Komputama Majenang¹, Universitas AMIKOM Yogyakarta²
bagus100486@gmail.com¹, emma@nrar.net²

INTISARI

Keluarga sejahtera adalah keluarga yang terbentuk dalam perkawinan sah, kebutuhan spiritual dan materiil yang layak, bertaqwa kepada Tuhan YME, mempunyai hubungan yang serasi, selaras dan seimbang dengan masyarakat dan lingkungan. Pemerintah telah melaksanakan berbagai program pembangunan keluarga sejahtera. Untuk mendukung hal tersebut, setiap tahun pemerintah melaksanakan proses pendataan keluarga. Pendataan keluarga dinilai sebagai langkah penting karena memiliki banyak fungsi, utamanya untuk memahami kelompok sasaran dan menentukan solusi untuk menyelesaikan masalah dari setiap kelompok sasaran itu. Dalam penelitian ini data mining digunakan untuk penentuan status tahapan keluarga sejahtera. Metode K-Nearest Neighbor (KNN), metode Naive Bayes dan Principal Component Analysis (PCA) digunakan untuk klasifikasi Penentuan status Tahapan keluarga sejahtera dengan tepat. Berdasarkan hasil pengujian, uji kinerja algoritma klasifikasi untuk kasus penentuan status tahapan keluarga sejahtera Kecamatan Sidareja untuk metode Naive Bayes menggunakan confusion matrix diperoleh akurasi 98,12% setelah ditambahkan seleksi fitur PCA menjadi 97,73% sedangkan metode KNN diperoleh akurasi sebesar 98,86%, kemudian setelah ditambahkan seleksi fitur PCA meningkat menjadi 98,96%

Kata kunci: *Keluarga Sejahtera, Data mining, K-Nearest Neighbor (KNN), Naive Bayes dan Principal Component Analysis (PCA)*

ABSTRACT

Prosperous family is a family formed in legitimate marriage, spiritual and material needs that are feasible, devoted to God YME, have a harmonious relationship, in harmony and balance with the community and the environment, the Government has implemented various development programs of prosperous families. To support this, every year the government implements family data collection process. Family data collection is considered an important step because it has many functions, primarily to understand the target group and to determine solutions to solve the problems of each target group. In this study data mining is used for the determination of the status of prosperous family stages. The K-Nearest Neighbor (KNN) method, the Naive Bayes method and the Principal Component Analysis (PCA) are used for the proper classification of status stages. Based on the results of the test, the performance of classification algorithm for the case of determining the status of prosperous family stages of Sidareja District for Naive Bayes method using confusion matrix obtained 98.12% accuracy after added PCA feature selection to 97.73% while KNN method obtained accuracy of 98.86% , then after added PCA feature selection increased to 98.96%

Keywords: *Prosperous Family, Data mining, K-Nearest Neighbor (KNN), Naive Bayes and Principal Component Analysis (PCA)*

I. PENDAHULUAN

Sejahtera merupakan kondisi atau keadaan yang baik, kondisi dimana manusia dalam keadaan makmur, sehat dan damai. Menurut Kamus Besar Bahasa Indonesia kata kesejahteraan merupakan kata benda yang mempunyai arti hal atau keadaan sejahtera, keamanan, keselamatan dan ketentraman. Sedangkan kata sejahtera sendiri memiliki arti aman sentosa dan makmur, selamat (terlepas

dari segala macam gangguan). Kesejahteraan dalam pada intinya mencakup dua hal pokok yaitu kesejahteraan sosial yang bersifat jasmani (lahir) dan rohani (batin). Sejahtera lahir dan batin tersebut harus terwujud dalam setiap pribadi (individu) yang bekerja untuk kesejahteraan hidupnya sendiri, sehingga akan terbentuk keluarga atau masyarakat dan negeri yang sejahtera.

Dengan disahkannya Undang-

Undang Nomor 10 Tahun 1992 tentang perkembangan Kependudukan dan Keluarga Sejahtera, misi Program Keluarga Berencana semakin luas. Pengertian Keluarga Berencana menjadi suatu upaya peningkatan kepedulian dan peran serta masyarakat melalui pendewasaan usia perkawinan, pengaturan kehamilan, pembinaan ketahanan keluarga dan peningkatan kesejahteraan keluarga kecil, bahagia dan sejahtera. [2]

Keluarga Sejahtera menurut UU No 10 tahun 1992 adalah keluarga yang dibentuk atas perkawinan yang sah, mampu memenuhi kebutuhan hidup spiritual dan materiil yang layak, bertaqwa pada Tuhan YME, memiliki hubungan yang serasi, selaras, dan seimbang antar anggota dan antara keluarga dengan masyarakat dan lingkungan [3]

Pendataan keluarga dinilai sebagai langkah penting karena memiliki banyak fungsi, utamanya untuk memahami kelompok sasaran dan menentukan solusi untuk menyelesaikan masalah dari setiap kelompok sasaran itu. Hal itu terjadi bila data yang diperoleh memiliki kualitas yang baik. Mengingat kualitas data menjadi isu yang penting. Proses pendataan keluarga sejahtera selama ini masih dilakukan secara manual atau secara langsung dengan melakukan kunjungan dari rumah ke rumah oleh petugas pendataan daerah. Hal ini berakibat pada pemborosan sumber daya manusia yang dimiliki sehingga potensi penurunan ketepatan dalam penentuan status tahapan keluarga sejahtera sangat dimungkinkan terjadi. Secara teoritis penentuan status tahapan keluarga sejahtera berdasarkan pemenuhan 21 indikator keluarga sejahtera. Akan tetapi untuk beberapa kasus, proses penentuan status tahapan keluarga sejahtera tidak lepas dari analisa para ahli / pakar sehingga menimbulkan banyak persepsi. [3].

Pada tahun 2015, Kecamatan Sidareja telah melakukan pendataan keluarga dengan menggunakan formulir F/I/MDK/08. Instrumen fomulir Pemutakhiran Data Keluarga (MDK) atau F/I/MDK/08, memuat variabel data keluarga yang lengkap dalam satu lembar untuk setiap keluarga. Dengan formulir MDK ini data keluarga dikumpulkan oleh para kader pendata bersama PLKB/PKB, lalu dihimpun di tingkat Kabupaten/Kota untuk selanjutnya dilakukan perekaman dan pengolahan. Pengumpulan data sub bidang KS dibantu PLKB (Petugas Lapangan Keluarga Berencana) kecamatan untuk melakukan sosialisasi secara langsung Melalui kantor dinas KB, PP, PA. Setiap data yang diperoleh

dalam kegiatan akan menunjukkan tingkat kesejahteraan keluarga.

Dalam kasus ini data mining dapat digunakan untuk menentukan status tahapan keluarga sejahtera. Data mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam database. Data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar [6]

Penelitian sebelumnya Untuk meningkatkan kecepatan waktu dan persentase akurasi dari metode penambangan data menggunakan metode KNN, Naïve Bayes dan kombinasi metode KNN, Naïve Bayes. Hasil keakuratan yang diperoleh dengan metode kombinasi KNN dan Naive Bayes, lebih tinggi dibanding dengan metode kombinasi Naive Bayes. Dilihat dari segi waktu, lama proses metode kombinasi KNN dan Naive Bayes lebih cepat dibanding metode KNN [17].

Pada penelitian ini penulis akan membandingkan dua metode dengan penambahan seleksi fitur dalam teknik data mining yaitu metode *K-Nearest Neighbor* (KNN), metode *Naive Bayes* dan *Principal Component Analysis (PCA)* untuk mendapatkan pengujian paling akurat dalam mengolah data dalam menentukan status tahapan keluarga sejahtera Kecamatan Sidareja.

II. METODE PENELITIAN

Metode penelitian yang digunakan adalah metode analisis deskriptif dengan pendekatan kuantitatif artinya penelitian yang dilakukan adalah menekankan analisisnya pada data-data numeric (angka), yang bertujuan untuk mendapatkan gambaran yang jelas mengenai suatu keadaan berdasarkan data yang diperoleh dengan cara menyajikan, mengumpulkan dan menganalisis data tersebut sehingga menjadi informasi baru yang dapat digunakan untuk menganalisa mengenai masalah yang sedang diteliti. Kemudian data – data tersebut akan diolah menggunakan data mining, Data mining sebenarnya merupakan salah satu bagian proses Knowledge Discovery in Database (KDD) yang bertugas untuk mengekstrak pola atau model dari data dengan menggunakan suatu algoritma yang spesifik. Adapun tahapan KDD sebagai berikut :

1. Data Selection
Pemilihan (seleksi) data dari sekumpulan data

operasional perlu dilakukan sebelum tahap penggalian informasi dalam knowledge data discovery (KDD) dimulai. Data hasil seleksi yang akan digunakan untuk proses data mining, disimpan dalam suatu berkas, terpisah dari basis data operasional.

2. Preprocessing atau Cleaning

Sebelum proses data mining dapat dilaksanakan, perlu dilakukan proses cleaning pada data yang menjadi fokus knowledge data discovery. Proses cleaning mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak juga dilakukan proses enrichment, yaitu proses memperkaya data yang sudah ada dengan data atau informasi lain yang relevan dan diperlukan untuk KDD, seperti data atau informasi.

3. Transformation

Coding adalah proses transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses data mining. Proses coding dalam knowledge data discovery merupakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam basis data.

4. Data mining

Data mining adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma dalam data mining sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat tergantung pada tujuan dan proses KDD secara keseluruhan.

5. Interpretation atau evaluation

Pola informasi yang dihasilkan dari proses data mining perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap ini merupakan bagian dari proses KDD yang disebut interpretation. Tahap ini mencakup pemeriksaan apakah pola informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada pada sebelumnya.

Tahap selanjutnya untuk mengukur akurasi algoritma klasifikasi, metode yang dapat digunakan antara lain cross validation, confusion matrix, dan kurva ROC (Receiver Operating Characteristic).

a. Cross Validation

Cross validation adalah pengujian standar yang dilakukan untuk memprediksi error rate. Data training dibagi secara random ke dalam beberapa 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.

b. Confusion matrix

Metode ini menggunakan tabel matriks seperti pada Tabel 1 jika data set hanya terdiri dari dua kelas, kelas yang satu dianggap sebagai positif dan yang lainnya negatif (Bramer, 2007).

TABEL I.

MODEL CONFUSION MATRIX

Klasifikasi yang benar	Diklasifikasikan sebagai	
	+	-
+	True Positives	False Negatives
-	False Positives	True Negatives

True positives adalah jumlah record positif yang diklasifikasikan sebagai positif, false positives adalah jumlah record negatif yang diklasifikasikan sebagai positif, false negatives adalah jumlah record positif yang diklasifikasikan sebagai negatif, true negatives adalah jumlah record negatif yang diklasifikasikan sebagai negative, kemudian masukkan data uji. Setelah data uji dimasukkan ke dalam confusion matrix, hitung nilai-nilai yang telah dimasukkan tersebut untuk dihitung jumlah sensitivity (recall), specificity, precision dan accuracy. Sensitivity digunakan untuk membandingkan jumlah TP terhadap jumlah record yang positif sedangkan specificity adalah perbandingan jumlah TN terhadap jumlah record yang negatif. Untuk menghitung digunakan persamaan di bawah ini (Han, 2006) :

$$\frac{TP}{P} = \frac{TP}{TP+FN} \dots \dots \dots (1)$$

Recall, True Positive Rate (Sensitivity)

Recall, True Negative Rate (Specificity)

$$= \frac{TN}{N} = \frac{TN}{TN+FP} \dots \dots \dots (2)$$

Precision, Positive Predictive Value

$$= \frac{TP}{TP+FP} \dots \dots \dots (3)$$

Precision, Negative Predictive Value

$$= \frac{TN}{TN+FN} \dots \dots \dots (4)$$

Accuracy =

$$\frac{TP+TN}{TP+FP+TN+FN} \dots \dots \dots (5)$$

Ukuran tingkat kesalahan klasifikasi juga dapat dihitung dengan mencari *Error Rate*:

$$Error\ Rate = \frac{FP+FN}{P+N} \dots \dots \dots (6)$$

Keterangan :

- TP : jumlah *true positif*
- TN : jumlah *true negatif*
- P : jumlah *record positif*
- N : jumlah *tupel negatif*
- FP : jumlah *false positif*
- N : jumlah *false negative*

III. HASIL DAN PEMBAHASAN

Langkah pertama diawali dengan pengumpulan data-data keluarga sejahtera yang didapatkan dari petugas lapangan untuk penentuan atribut-atribut yang akan

digunakan. Data atribut Keluarga Sejahtera dapat dilihat pada Tabel .2

Selanjutnya *Data Cleaning* (Pembersihan Data) Pembersihan data pada penelitian ini dilakukan secara manual, dikarenakan data mentah berupa kertas cetak, sehingga proses pembersihan data dilakukan diluar aplikasi, Data dibersihkan dari beberapa item yang tidak memiliki *value*, Hanya data yang memiliki atribut yang lengkap dan memenuhi persyaratan kebutuhan yang dipilih.

Transformasi data digunakan untuk mengubah dataset sehingga konten informasi terbaik diambil dan dengan melakukan pengurangan atau perubahan tipe data *standart* sehingga data siap digunakan untuk dipresentasikan ke teknik data mining menggunakan *RapidMiner*, Proses klasifikasi diawali dengan penentuan dataset yang disimpan dalam format excel(*,xls), Tabel 3 merupakan pembagian variabel dan kelas data yang digunakan dalam analisis data mining, Sedangkan Tabel 4 menunjukkan sampel dari dataset yang digunakan.

TABEL II.
ATRIBUT KELUARGA SEJAHTERA

NO	Atribut	Keterangan
1	Keluarga	Merupakan Nama Keluarga
2	Indikator 1 (A)	Pada umumnya anggota keluarga makan dua kali sehari atau lebih
3	Indikator 2 (B)	Anggota keluarga memiliki pakaian yang berbeda untuk di rumah, bekerja / sekolah dan bepergian
4	Indikator 3 (C)	Rumah yang ditempati keluarga mempunyai atap, lantai, dinding yang baik
5	Indikator 4 (D)	Bila ada anggota keluarga sakit dibawa ke sarana kesehatan
6	Indikator 5 (E)	Bila pasangan usia subur ingin ber KB pergi ke sarana pelayanan kontrasepsi
7	Indikator 6 (F)	Semua anak umur 7-15 tahun dalam keluarga bersekolah
8	Indikator 7 (G)	Pada umumnya anggota keluarga melaksanakan ibadah sesuai dengan agama dan kepercayaan masing-masing
9	Indikator 8 (H)	Paling kurang sekali seminggu seluruh anggota keluarga makan daging / ikan / telur
10	Indikator 9 (I)	Seluruh anggota keluarga memperoleh paling kurang satu pasang pakaian baru dalam setahun
11	Indikator 10 (J)	Luas lantai rumah paling kurang 8 m2 untuk setiap penghun i rumah
12	Indikator 11 (K)	Tiga bulan terakhir keluarga dalam keadaan sehat, sehingga dapat melaksanakan tugas/fungsi masing-masing
13	Indikator 12 (L)	Ada seorang atau lebih anggota keluarga yang bekerja untuk memperoleh penghasilan
14	Indikator 13 (M)	Seluruh anggota keluarga umur 10 - 60 tahun bisa baca tulisan latin
15	Indikator 14 (N)	Pasangan usia subur dengan anak dua atau lebih menggunakan alat/obat kontrasepsi
16	Indikator 15 (O)	Keluarga berupaya meningkatkan pengetahuan agama
17	Indikator 16 (P)	Sebagian penghasilan keluarga ditabung dalam bentuk uang atau barang

18	Indikator 17 (Q)	Kebiasaan keluarga makan bersama paling kurang seminggu sekali dimanfaatkan untuk berkomunikasi
19	Indikator 18 (R)	Keluarga ikut dalam kegiatan masyarakat di lingkungan tempat tinggal
20	Indikator 19 (S)	Keluarga memperoleh informasi dari surat kabar / majalah / radio / tv
21	Indikator 20 (T)	Keluarga secara teratur dengan suka rela memberikan sumbangan materiil untuk kegiatan social
22	Indikator 21 (U)	Ada anggota keluarga yang aktif sebagai pengurus perkumpulan sosial/yayasan/ institusi masyarakat
22	Status (class)	Merupakan status keluarga terdiri dari : Pra Sejahtera; KS I; KS II; KS III; KS III +

TABEL III.
PEMBAGIAN VARIABEL DAN KELAS DATA

Variabel	Nama Field	Jenis Data	Kelas Data yang digunakan
Class	Status	Polynomial	Pra Sejahtera; KS I; KS II; KS III; KS III +
A	KS I	Binomial	1 ; 0
B		Binomial	1 ; 0
C		Binomial	1 ; 0
D		Binomial	1 ; 0
E		Binomial	1 ; 0
F		Binomial	1 ; 0
G	KS II	Binomial	1 ; 0
H		Binomial	1 ; 0
I		Binomial	1 ; 0
J		Binomial	1 ; 0
K		Binomial	1 ; 0
L		Binomial	1 ; 0
M	KS III	Binomial	1 ; 0
N		Binomial	1 ; 0
O		Binomial	1 ; 0
P		Binomial	1 ; 0
Q		Binomial	1 ; 0
R	KS III +	Binomial	1 ; 0
S		Binomial	1 ; 0
T		Binomial	1 ; 0
U		Binomial	1 ; 0

Nilai kelas pada field Klasifikasi (Y) dikategorikan berdasarkan kriteria berikut ini :

- 1) Tahapan Pra Sejahtera, keluarga yang belum dapat memenuhi salah satu indikator tahapan Keluarga Sejahtera I
- 2) Tahapan Pra Sejahtera, keluarga yang belum dapat memenuhi salah satu indikator tahapan Keluarga Sejahtera I
- 3) Tahapan Keluarga Sejahtera I, keluarga yang baru dapat memenuhi indikator-indikator 1 s/d 6.
- 4) Tahapan Keluarga Sejahtera II, keluarga yang sudah dapat memenuhi indikator Tahapan Keluarga Sejahtera I (indikator 1 s/d 6) dan indikator 7 s/d 14.
- 5) Tahapan Keluarga Sejahtera III, Adalah keluarga yang sudah memenuhi indikator

- 6) Tahapan keluarga Sejahtera I dan Indikator Keluarga Sejahtera II (Indikator 1 s/d 14) dan indikator 15 s/d 19
Tahapan Keluarga Sejahtera III Plus(+), keluarga yang memenuhi indikator Tahapan keluarga Sejahtera I, Indikator Keluarga Sejahtera II dan Indikator Keluarga Sejahtera III (Indikator 1 s/d 19) dan indikator 20 s/d 21.

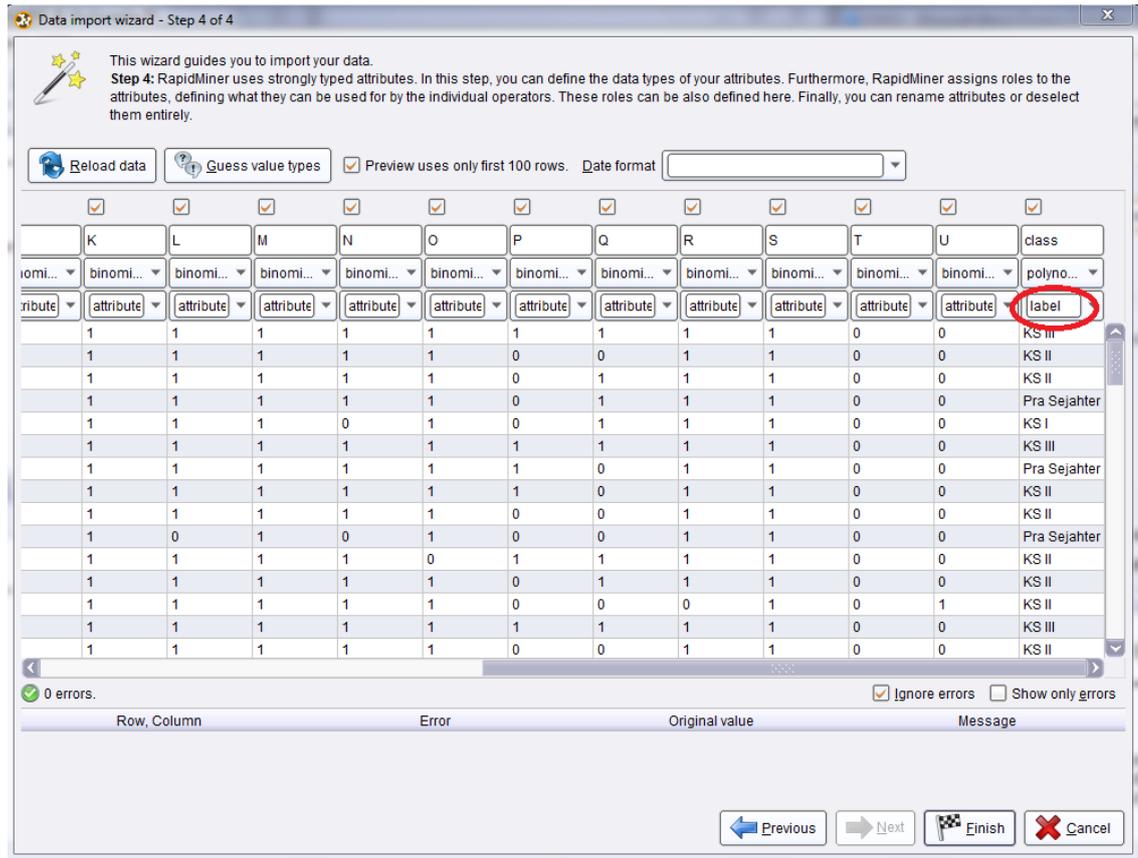
Nilai kelas pada field 1 s/d 21 (X) dikategorikan berdasarkan keluarga yang memenuhi syarat indikator 1 s/d 21, jika memenuhi diberikan angka satu (1) dan jika tidak memenuhi diberikan angka nol (0), Pendataan data keluarga dapat dilihat pada table 4.

TABEL IV.
DATASET DATA KELUARGA

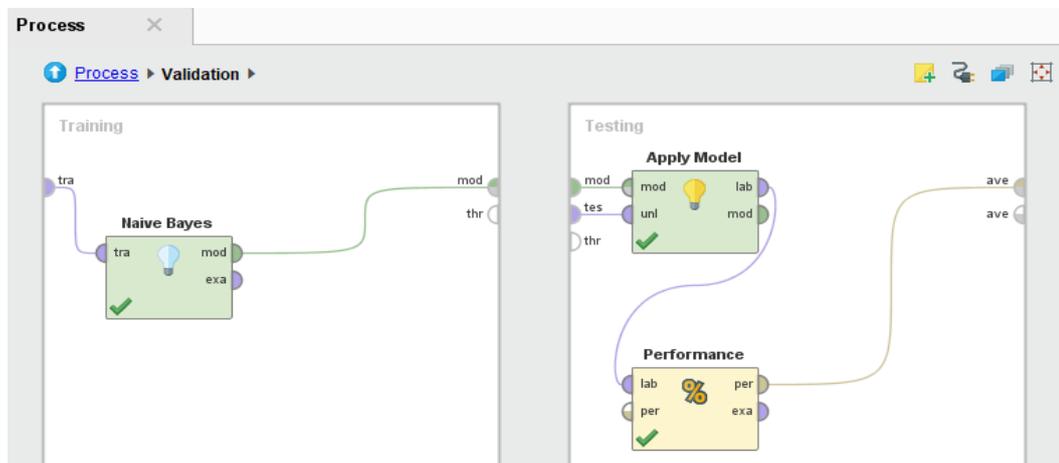
A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	class
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	KS III
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	0	0	KS II
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	0	0	KS II
1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	0	1	1	1	0	0	Pra Sejahtera
1	1	1	1	1	1	1	0	1	1	1	1	1	0	1	0	1	1	1	0	0	KS I
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	KS III
1	1	1	1	0	0	1	1	1	1	1	1	1	1	1	0	1	1	0	0	0	Pra Sejahtera
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0	0	0	KS II
1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	0	0	1	1	0	0	KS II
1	1	0	1	1	1	1	0	1	1	1	0	1	0	1	0	0	1	1	0	0	Pra Sejahtera
1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	0	0	KS II
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	0	0	KS II
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	0	1	KS II
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	KS III
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	0	0	KS II
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	0	0	KS II
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	0	0	KS II

Dalam penentuan dataset pada RapidMiner, Atribut yang menjadi tujuan diisi dalam proses data mining merupakan atribut target sehingga menurut aturan atribut tersebut harus diganti menjadi 'label' ditunjukkan pada Gambar 1. Pada tahap selanjutnya yaitu tahap klasifikasi menggunakan *RapidMiner*, dataset yang telah di-import untuk selanjutnya dilakukan proses *training dan testing* dengan algoritma yang diusulkan, yaitu pertama

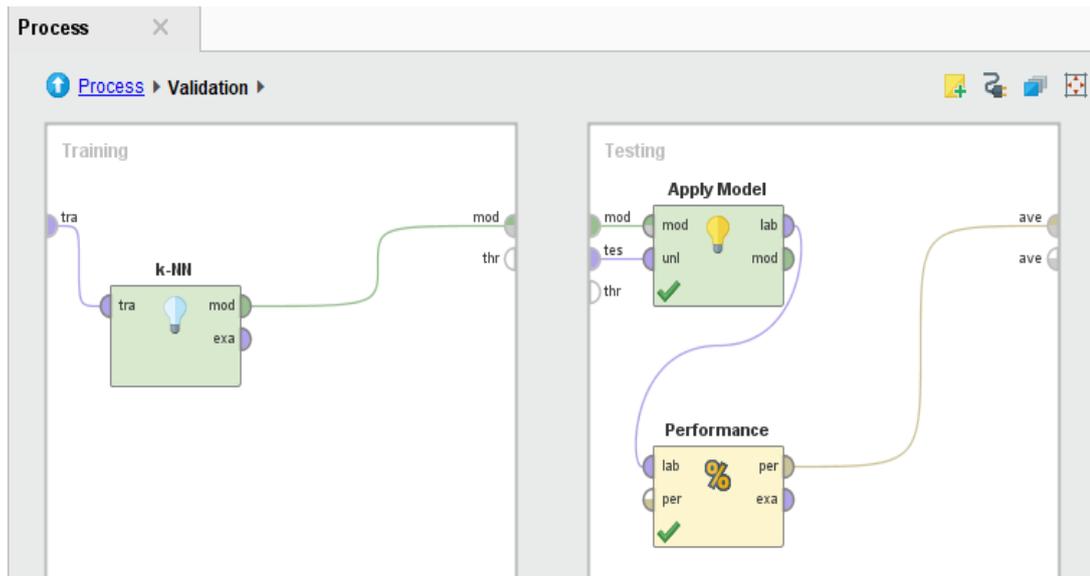
menggunakan *NaiveBayes* dan yang kedua algoritma KNN. Didalam kolom *training* terdapat algoritma klasifikasi yang diterapkan, yaitu algoritma *NaiveBayes* seperti pada gambar 2 dan algoritma KNN pada gambar 3 Sedangkan didalam kolom *testing* terdapat *Apply Model* untuk menjalankan model algoritma yang dipilih serta *Performance* untuk mengukur performa / kinerja dari masing-masing model algoritma.



Gambar 1. Atribut Variabel Target Dalam Format 'label'.



Gambar 2. Proses testing dan training algoritma NaiveBayes



Gambar 3. Proses testing dan training algoritma KNN

Proses seleksi variabel data set menggunakan metode *Principal Component Analysis (PCA)* dengan cara mencocokkan variabel/atribut yang awal dengan variabel/atribut yang ada pada

basis kasus. Hasil seleksi fitur berdasarkan PCA dapat dilihat pada Tabel 5 Pada masing-masing metode seleksi fitur menghasilkan fitur yang berbeda.

TABEL V.
HASIL SELEKSI FITUR BERDASARKAN PCA

Metode Seleksi Fitur	Atribut/fitur
PCA	Fitur 1, Fitur 2, Fitur 3, Fitur 4, Fitur 5, Fitur 6, Fitur 7, Fitur 8, Fitur 9, Fitur 10, Fitur 11, Fitur 12, Fitur 13

Berdasarkan Tabel 4 jumlah fitur yang terseleksi dengan metode PCA sebanyak 13 atribut atau fitur yang terseleksi, Metode PCA akan mentransformasi fitur pada dataset kebentuk fitur baru, sehingga pada penelitian ini diberi nama Fitur 1, Fitur 2, Fitur 3, Fitur 4, Fitur 5, Fitur 6, Fitur 7, Fitur 8, Fitur 9, Fitur 10, Fitur 11, Fitur 12, Fitur 13

Atribut atau fitur yang terseleksi akan digunakan pada proses selanjutnya yaitu proses klasifikasi, sedangkan fitur yang tidak terseleksi akan dihapus beserta datanya dari dataset, Sehingga terdapat 8 atribut beserta datanya yang dihapus dari dataset yang digunakan.

Penambahan seleksi fitur *Principal Component Analysis (PCA)* dengan algoritma *Naive Bayes* dan *Principal Component Analysis (PCA)* dengan algoritma KNN. Tahap selanjutnya yaitu untuk mengetahui tingkat akurasi Algoritma klasifikasi yang digunakan

yaitu *Naive Bayes* dan KNN, tanpa menggunakan penambahan seleksi fitur algoritma tersebut akan menghasilkan *performa* yang menjadi tolak ukur dalam proses klasifikasi, Akurasi yang dihasilkan algoritma *Naive bayes* mencapai 95,31% sedangkan pada algoritma KNN mencapai 97,28% Dan penambahan seleksi fitur menggunakan PCA yaitu algoritma *Naive Bayes* dengan seleksi fitur PCA dan KNN dengan seleksi fitur PCA mencapai tingkat akurasi sebesar 94,32% untuk *Naive bayes* dengan PCA dan KNN dengan seleksi fitur PCA sebesar 97,41% seperti gambar 4.

accuracy: 95.31% +/- 1.73% (mikro: 95.31%) Bayes						
	true KS III	true KS II	true Pra Sejahtera	true KS I	true KS III +	class precision
pred. KS III	145	0	0	0	0	100.00%
pred. KS II	3	135	0	1	0	97.12%
pred. Pra Sejahtera	0	0	296	3	1	98.67%
pred. KS I	0	30	0	196	0	86.73%
pred. KS III +	0	0	0	0	0	0.00%
class recall	97.97%	81.82%	100.00%	98.00%	0.00%	

accuracy: 94.32% +/- 3.54% (mikro: 94.32%) Bayes + PCA						
	true KS III	true KS II	true Pra Sejahtera	true KS I	true KS III +	class precision
pred. KS III	130	1	0	0	0	99.24%
pred. KS II	3	148	0	1	0	97.37%
pred. Pra Sejahtera	15	9	296	9	1	89.70%
pred. KS I	0	7	0	190	0	96.45%
pred. KS III +	0	0	0	0	0	0.00%
class recall	87.84%	89.70%	100.00%	95.00%	0.00%	

accuracy: 97.28% +/- 0.84% (mikro: 97.28%) KNN						
	true KS III	true KS II	true Pra Sejahtera	true KS I	true KS III +	class precision
pred. KS III	145	0	1	0	1	98.64%
pred. KS II	3	158	0	10	0	92.40%
pred. Pra Sejahtera	0	0	295	0	0	100.00%
pred. KS I	0	7	0	190	0	96.45%
pred. KS III +	0	0	0	0	0	0.00%
class recall	97.97%	95.76%	99.66%	95.00%	0.00%	

accuracy: 97.41% +/- 1.51% (mikro: 97.41%) KNN + PCA						
	true KS III	true KS II	true Pra Sejahtera	true KS I	true KS III +	class precision
pred. KS III	145	0	0	0	1	99.32%
pred. KS II	3	158	0	10	0	92.40%
pred. Pra Sejahtera	0	0	296	0	0	100.00%
pred. KS I	0	7	0	190	0	96.45%
pred. KS III +	0	0	0	0	0	0.00%
class recall	97.97%	95.76%	100.00%	95.00%	0.00%	

Gambar 4. Hasil Uji Coba Algoritma Naive Bayes, KNN dan PCA

Tahap Selanjutnya mencari nilai *confusion matrix* untuk setiap jenis/level untuk model yang menggunakan *Naive Bayes*, KNN, PCA + *Naive Bayes*, dan PCA + KNN berdasarkan hasil uji coba algoritma. Kedua, dari hasil *confusion matrix* setiap jenis/tingkat, kemudian mengukur kinerja berdasarkan akurasi, *sensitivity*, *specificity*, dan *precision* (PPV). Berdasarkan *confusion matrix* dengan algoritma *Naive Bayes* yang didapatkan, akan diperoleh nilai TP, TN, FP, dan FN untuk masing-masing kelas, Berikut merupakan nilai untuk kelas a (KS I):

1. True Positif (TP) = 196
2. True Negatif (TN) = 135 + 3 + 0 + 0 + 0 + 145 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 1 + 196 = 580
3. False Positif (FP) = 1 + 0 + 0 + 3 = 4
4. False Negatif (FN) = 30 + 0 + 0 + 0 = 30

Berikut merupakan nilai TP, TN, FP, dan FN untuk kelas b (KS II):

1. True Positif (TP) = 135
2. True Negatif (TN) = 145 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 1 + 296 + 3 + 196 + 0 + 0 + 0 = 641

3. False Positif (FP) = 0 + 0 + 0 + 30 = 30

False Negatif (FN) = 3 + 0 + 0 + 1 = 4
 Berikut merupakan nilai TP, TN, FP, dan FN untuk kelas c (KS III):

True Positif (TP) = 145
 True Negatif (TN) = 0 + 0 + 0 + 0 + 1 + 296 + 3 + 0 + 196 + 30 + 0 + 0 + 1 + 135 + 0 + 0 = 662

False Positif (FP) = 0 + 0 + 0 + 3 = 3
 False Negatif (FN) = 0 + 0 + 0 + 0 = 0

Berikut merupakan nilai TP, TN, FP, dan FN untuk kelas d (KS III +):

True Positif (TP) = 0
 True Negatif (TN) = 296 + 3 + 0 + 0 + 196 + 30 + 0 + 0 + 1 + 135 + 3 + 0 + 0 + 0 + 145 + 0 = 809

False Positif (FP) = 1 + 0 + 0 + 0 = 1
 False Negatif (FN) = 0 + 0 + 0 + 0 = 0

Berikut merupakan nilai TP, TN, FP, dan FN untuk kelas e (Pra Sejahtera):

1. True Positif (TP) = 296
 2. True Negatif (TN) = 196 + 30 + 0 + 0 + 1 + 135 + 3 + 0 + 0 + 0 + 145 + 0 + 0 + 0 + 0 + 0 = 510

3. False Positif (FP) = 0 + 0 + 0 + 0 = 0
 4. False Negatif (FN) = 3 + 0 + 0 + 1 = 4

Berdasarkan confusion matrix dengan algoritma KNN yang didapatkan, akan diperoleh nilai TP, TN, FP, dan FN untuk masing-masing kelas, Berikut merupakan nilai untuk kelas a (KS I):

1. True Positif (TP) = 190
2. True Negatif (TN) = 158 + 3 + 0 + 0 + 0 + 145 + 1 + 1 + 0 + 0 + 0 + 0 + 0 + 1 + 0 + 295 = 603
3. False Positif (FP) = 10 + 0 + 0 + 0 = 10
4. False Negatif (FN) = 7 + 0 + 0 + 0 = 7

Berikut merupakan nilai TP, TN, FP, dan FN untuk kelas b (KS II):

1. True Positif (TP) = 158
2. True Negatif (TN) = 145 + 1 + 1 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 295 + 0 + 190 + 0 + 0 + 0 = 632
3. False Positif (FP) = 7 + 0 + 0 + 0 = 7
4. False Negatif (FN) = 3 + 0 + 0 + 10 = 13

Berikut merupakan nilai TP, TN, FP, dan FN untuk kelas c (KS III):

1. True Positif (TP) = 145
2. True Negatif (TN) = 0 + 0 + 0 + 0 + 0 + 295 + 0 + 0 + 190 + 7 + 0 + 0 + 0 + 10 + 158 + 0 + 0 = 660

False Positif (FP) = 0 + 0 + 3 + 0 = 3
 False Negatif (FN) = 1 + 1 + 0 + 0 = 2
 Berikut merupakan nilai TP, TN, FP, dan FN untuk kelas d (KS III +):

1. True Positif (TP) = 0
 2. True Negatif (TN) = 295 + 0 + 0 + 0 + 190 + 7 + 0 + 0 + 10 + 158 + 3 + 0 + 0 + 0 + 0 + 145 + 1 = 809
 3. False Positif (FP) = 0 + 0 + 0 + 1 = 1
 4. False Negatif (FN) = 0 + 0 + 0 + 0 = 0
- Berikut merupakan nilai TP, TN, FP, dan FN untuk kelas e (Pra Sejahtera):

1. True Positif (TP) = 295
2. True Negatif (TN) = 190 + 7 + 0 + 0 + 10 + 158 + 3 + 0 + 0 + 0 + 145 + 1 + 0 + 0 + 0 + 0 = 514
3. False Positif (FP) = 0 + 1 + 0 + 0 = 1
4. False Negatif (FN) = 0 + 0 + 0 + 0 = 0

Dari data uji yang sudah didapat, selanjutnya menghitung nilai kinerja *sensitivity* (*recall*), *specificity*, *precision* *Positive Predictive Value*, *Negative Predictive Value* dan *accuracy*. Perhitungan nilai *Sensitivity*, *Specifity*, *PPV*, *NPV* dan *Accuracy* diperoleh pada tabel 6.

TABEL VI.
 NILAI *SENSITIVITY*, *SPECIFITY*, *PPV*, *NPV* DAN *ACCURACY*

Kelas	Metode	<i>Sensitivity</i>	<i>Specifity</i>	PPV	NPV	<i>Accuracy</i>
a=KS I	Algoritma <i>Naive Bayes</i>	86,73 %	99,32 %	98,00 %	95,08 %	95,80 %
	Algoritma KNN	96,42 %	98,20 %	94,5 %	98,85 %	97,77 %
	Algoritma <i>Naive Bayes</i> +PCA	96,45 %	98,37 %	95,00 %	98,85 %	97,90 %
	Algoritma KNN + PCA	96,45 %	98,37 %	95,00 %	98,85 %	97,90 %
b=KS II	Algoritma <i>Naive Bayes</i>	97,12 %	95,53 %	81,82 %	99,38 %	95,80 %
	Algoritma KNN	91,86 %	98,90 %	95,76 %	97,98 %	97,4 %
	Algoritma <i>Naive Bayes</i> +PCA	97,37 %	97,42 %	89,70 %	99,38 %	97,41 %
	Algoritma KNN + PCA	92,40 %	98,90 %	95,76 %	97,98 %	97,53 %
c=KS III	Algoritma <i>Naive Bayes</i>	100,00 %	99,55 %	97,97 %	100,00 %	99,63 %
	Algoritma KNN	98,64 %	99,55 %	97,97 %	99,70 %	99,38 %
	Algoritma <i>Naive Bayes</i> +PCA	99,24 %	97,35 %	87,84 %	99,85 %	97,65 %
	Algoritma KNN + PCA	99,32 %	99,55 %	97,97 %	99,85 %	99,51 %
d=KS III +	Algoritma <i>Naive Bayes</i>	0,00 %	99,88 %	0,00 %	100,00 %	99,88 %
	Algoritma KNN	0,00 %	99,88 %	0,00 %	100,00 %	99,88 %
	Algoritma <i>Naive Bayes</i> +PCA	0,00 %	99,88 %	0,00 %	100,00 %	99,88 %
	Algoritma KNN + PCA	0,00 %	99,88 %	0,00 %	100,00 %	99,88 %

e=Pra Sejahtera	Algoritma <i>Naive Bayes</i>	98,67 %	100,00 %	100,00 %	99,22 %	99,51 %
	Algoritma KNN	100,00 %	99,81 %	99,66 %	100,00 %	99,88 %
	Algoritma <i>Naive Bayes</i> +PCA	89,70 %	100,00 %	100,00 %	93,39 %	95,80 %
	Algoritma KNN + PCA	100,00 %	100,00 %	100,00 %	100,00 %	100,00 %
Average	Algoritma <i>Naive Bayes</i>	76,50 %	98,85 %	75,56 %	98,74 %	98,12 %
	Algoritma KNN	77,38 %	99,27 %	77,58 %	99,27 %	98,86 %
	Algoritma <i>Naive Bayes</i> +PCA	76,55 %	98,60 %	74,51 %	98,29 %	97,73 %
	Algoritma KNN + PCA	77,63 %	99,34 %	77,75 %	99,34 %	98,96 %

IV. KESIMPULAN

Berdasarkan hasil implementasi KNN, *Naive Bayes* dan seleksi fitur PCA pada kasus penentuan tahapan keluarga sejahtera Kecamatan Sidareja dapat diambil beberapa kesimpulan sebagai berikut :

- Hasil uji kinerja algoritma klasifikasi untuk kasus penentuan status tahapan keluarga sejahtera untuk algoritma *Naive Bayes* menggunakan *confusion matrix* dengan $K = 10$ diperoleh akurasi 98,12% setelah ditambahkan seleksi fitur PCA menjadi 97,73% sedangkan algoritma KNN diperoleh akurasi sebesar 98,86%, kemudian setelah ditambahkan seleksi fitur PCA meningkat menjadi 98,96%.
- Dengan penambahan seleksi fitur PCA algoritma *Naive Bayes* tingkat akurasinya menurun dari 98,12% menjadi 97,73% sedangkan algoritma KNN ditambahkan seleksi fitur PCA tingkat akurasi meningkat dari 98,86% menjadi 98,96%. Jadi dengan penambahan seleksi fitur PCA tidak terlalu banyak mempengaruhi perhitungan akurasi.
- Kinerja algoritma KNN lebih unggul dibandingkan algoritma *Naive Bayes* pada kasus penentuan status tahapan keluarga sejahtera. Hal ini terlihat dari penggunaan *confusion matrix* dalam mengukur nilai kinerja berdasarkan *sensitivity (recall)*, *specificity*, *precision Positive Predictive Value*, *Negative Predictive Value* dan *accuracy* yang di peroleh masing-masing algoritma.
- Perhitungan akurasi algoritma KNN lebih unggul dibandingkan algoritma *Naive Bayes* pada kasus penentuan status tahapan keluarga sejahtera.
- Berdasarkan pengujian klasifikasi algoritma KNN memiliki kinerja yang baik dibandingkan *Naive Bayes*, berdasarkan *sensitivity (recall)*, *specificity*, *precision Positive Predictive Value*, *Negative Predictive Value* dan *accuracy*.

UCAPAN TERIMAKASIH

Penulis ucapkan kepada Bapak Rom

Muhdlori, S.Ag, M.M selaku Kepala UPT dan Nofi Sulistiyowati, S.Km selaku Kasubag TU UPT Dinas KB, PP, PA yang telah membantu dalam tempat penelitian.

REFERENSI

- Bramer, M., 2007, Principles of Data Mining, Springer, London.
- BKKBN, 2009, Pedoman Tata Cara Pencatatan dan Pelaporan Pendataan Keluarga Tahun 2007, BKKBN Provinsi Jawa Tengah
- BKKBN, 2012, Profil Hasil Pendataan Keluarga Tahun 2012, Direktorat Pelaporan dan Statistik. Jakarta.
- Florin Gorunescu, 2011 *Data Mining: Concepts, Models and Techniques*, Springer
- Han, J., & Kamber, M, 2006, *Data Mining Concept and Tehniques*. San Fransisco : Morgan Kauffman.
- Kusrini, & Luthfi. E T, 2009, *Algoritma Data Mining* Yogyakarta : Andi.
- Larose, D.T., 2005, *Discovering Knowledge in Data*. New Jersey, John Willey & Sons, Inc
- Larose, D T, 2006, *Data Mining Methods and Models*. Hoboken New Jersey : Jhon Wiley & Sons, Inc,
- Liao, 2007, *Recent Advances in Data Mining of Enterprise Data : Algorithms and Application*. Singapore : World Scientific Publishing
- Saaty, T.L., 1980, *The Analytic Hierarchy Process*, McGraw-Hill, New York
- Alfin Dhuhawan Bagja, Gunawan Abdillah & Faiza Renaldi, 2016, Penerapan Data Mining Dalam Menentukan Potensi Keberhasilan bakal Calon Legislatif Di Daerah Pemilihan Jawa Barat Menggunakan Algoritma *K-Nearest Neighbors*, Prosiding SNST, ISBN 978-602-99334-5-1

- [12] Burhan Alfironi Muktamar, Noor Akhmad Setiawan, Teguh Bharata Adji, 2015, Pembobotan Korelasi Pada Naïve Bayes Classifier, Seminar Nasional Teknologi Informasi dan Multimedia 2015, *ISSN : 2302-3805*
- [13] Fathur Rahman dan Muhammad Iqbal Firdaus, 2016, Penerapan Data Mining Metode Naïve Bayes Untuk Prediksi Hasil Belajar Siswa Sekolah Menengah Pertama (SMP), *Al Ulum Sains dan Teknologi Vol.1 No.2*
- [14] Muktamar et al, 2015, Pembobotan Korelasi Pada Naïve Bayes Classifier, Semnas Teknomedia, *ISSN : 2302-3806*
- [15] Mustakim dan Giantika Oktaviani, 2016, Algoritma *K-Nearest Neighbor Classification* Sebagai Sistem Prediksi Predikat Prestasi Mahasiswa, *Jurnal Sains, Teknologi dan Industri*, *ISSN : 1693-2390*
- [16] Ndaumanu, Ricky Imanuel., Kusri., Arief, M. Rudyanto., 2014. Analisis Prediksi Tingkat Pengunduran Diri Mahasiswa dengan Metode *K-Nearest Neighbor*. Jatsi, Vol. 1 No. 1
- [17] Sari M, 2015, Kombinasi Metode *K-Nearest Neighbor* Dan *Naïve Bayes* Untuk Klasifikasi Data, *ISSN : 2302-3806*
- [18] Soemartini., 2008, Principal Component Analisis sebagai Salah Satu untuk Mengatasi Multikolinearitas. Jatinangor:FMIPA-UNPAD
- [19] Susetyoko, R & Purwantini, E, 2009, Teknik Reduksi Dimensi Menggunakan Komponen Utama Data Partisi Pada Pengklasifikasian Data Berdimensi Tinggi dengan Ukuran Sampel Kecil. *Jurnal. Politeknik Elektronika Negeri Surabaya*
- [20] Wiharto W, M.Com et al, 2016, *Interpretation of Clinical Data Based On C4.5 Algorithm for the Diagnosis of Coronary Heart Disease*, *Healthcare Information Research*, *eISSN : 2093-369X*