



**KLASIFIKASI KESEJAHTERAAN RUMAH TANGGA DI  
KABUPATEN BREBES MENGGUNAKAN KOMBINASI  
*PRINCIPAL COMPONENT ANALYSIS* DAN *SUPPORT  
VECTOR MACHINE (PCA-SVM)***

**JURNAL ILMIAH**

**Diajukan sebagai salah satu syarat memperoleh gelar Sarjana**

**Oleh**

**IRMA ARIANTI**

**B2A016027**

**PROGRAM STUDI S1 STATISTIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS MUHAMMADIYAH SEMARANG**

**2020**

## HALAMAN PENGESAHAN

Skripsi dengan judul “Klasifikasi Kesejahteraan Rumah Tangga di Kabupaten Brebes Menggunakan Kombinasi *Principal Component Analysis* dan *Support Vector Machine* (PCA-SVM)” yang disusun oleh :

Nama : Irma Arianti

NIM : B2A016027

Program Studi : S1 Statistika

Telah disetujui oleh dosen pembimbing pada 5 Mei 2020.

Semarang, 17 April 2020

Pembimbing Utama	Pembimbing Pendamping
 Tiani Wahyu Utami, S.Si, M.Si	 Indah Manfaati Nur, S.Si, M.Si
NIK. 28.6.1026.341	NIK. 28.6.1026.221

Mengetahui,

Ketua Program Studi Statistika



Indah Manfaati Nur, S.Si, M.Si

NIK. 28.6.1026.221

# KLASIFIKASI KESEJAHTERAAN RUMAH TANGGA DI KABUPATEN BREBES MENGGUNAKAN *PRINCIPAL COMPONENT ANALYSIS* DAN *SUPPORT VECTOR MACHINE*

Irma Arianti, Tiani Wahyu Utami, dan Indah Manfaati Nur  
Program Studi S1 Statistika, Universitas Muhammadiyah Semarang  
irmaarianti4@gmail.com

## ABSTRAK

Welfare is a variety of actions taken by humans to achieve a better level of community life. Households who are unable to fulfill their basic needs are categorized as poverty. Based on PPLS 2008, there are 13 indicators in determining poverty obtained from the SUSENAS results. If there are 9 or more variables that meet the poverty criteria, then it is categorized as a poor household and vice versa. This study uses a combination of Principal Component Analysis (PCA) and Support Vector Machine (SVM) methods to classify household welfare in Brebes Regency in 2018 with the poor category. PCA is used to reduce data and the latest data is processed using SVM to be classified. Used in 3 categories of training and testing data sharing namely 70:30, 80:20 and 90:10. Classification using PCA-SVM in training data produces the best level of accuracy, namely the training data 80 of 99.87%. The greatest sensitivity is 100% in training data 70 and 80 while the best specificity in data 90 is 99.86%. For testing data has an accuracy and sensitivity level of 99.3% and 94.59%, namely the 30 testing data, and specificity has the best value on all data sharing by 100%.

**Keyword :** Household welfare, *Principal Component Analysis* (PCA), *Support Vector Machine* (SVM)

## Pendahuluan

Kemiskinan merupakan salah satu permasalahan yang sering dihadapi dalam peningkatan kesejahteraan di suatu Negara. Permasalahan kemiskinan ini merupakan kasus yang tidak lekang oleh waktu yang membutuhkan waktu lama untuk mengatasinya. Permasalahan ini sering dialami oleh Negara berkembang. Masyarakat miskin merupakan suatu kondisi dimana fisik masyarakat yang tidak memiliki akses sarana dan prasarana yang memadai dengan kualitas pemukiman yang dibawah standar kelayakan.

Fenomena kemiskinan juga dapat dilihat dari hubungan kausalitas yang menjelaskan sebab-sebab kejadian kemiskinan. Menurut Bambang Subagio *et all* (2001) yang dikutip dalam Institut

Management Zakat (IMZ), berdasarkan kajian data empiris menyebutkan bahwa sebab-sebab kemiskinan dapat dibagi menjadi dua golongan yaitu kemiskinan disebabkan oleh faktor alamiah dan non alamiah. Kemiskinan disebabkan karena faktor alamiah merupakan kemiskinan yang dilihat dari kondisi lingkungannya, pengetahuan yang kurang memadai dan faktor alam lainnya. Sedangkan kemiskinan disebabkan karena faktor non alamiah merupakan kemiskinan yang terjadi karena terdapat kesalahan kebijakan ekonomi, adanya korupsi yang merajalela, kondisi politik yang tidak stabil, kesalahan pengelolaan sumber daya alam dan faktor lainnya.

Di Indonesia, jumlah penduduk miskin tahun 2018 mencapai 25.7 juta jiwa atau sebesar 9.66% dari 260 juta

jiwa. Angka ini mengalami penurunan dari tahun sebelumnya. Di Jawa Tengah, angka kemiskinan mencapai 3.8 juta jiwa atau 11.32% dari 34.49 juta jiwa. Angka tersebut menunjukkan bahwa persentase penduduk miskin di Jawa Tengah lebih tinggi dari Indonesia. Dari data tersebut, maka data penduduk di Kabupaten Brebes akan digunakan dalam penelitian ini.

Untuk mengetahui jumlah penduduk miskin di suatu daerah, pemerintah perlu melakukan klasifikasi status kemiskinan suatu rumah tangga. Salah satu metode klasifikasi yang dapat digunakan adalah *Support Vector Machine* (SVM). SVM memiliki tujuan membangun sebuah model yang dapat memprediksi data tes yang diberikan. Metode SVM dapat digunakan dengan waktu pengujian yang singkat akan tetapi perlu memperkecil beban komputasinya.

Salah satu cara untuk memperkecil beban komputasi data sebelum dilakukan uji dengan menggunakan SVM adalah melakukan ekstraksi fitur. Ekstraksi fitur digunakan untuk mendapatkan ciri-ciri utama dari data. Metode yang sering digunakan dalam mengekstraksi fitur adalah *Principal Component Analysis* (PCA). PCA ini digunakan untuk mereduksi dimensi dari sebuah data. Pada penelitian ini, PCA dikombinasi dengan SVM untuk mengklasifikasikan kesejahteraan rumah tangga.

Penelitian yang dilakukan Abdillah dkk, tentang Pembelajaran Mesin Menggunakan *Principal Component Analysis* dan *Support Vector Machine* untuk Mendeteksi Diabetes menghasilkan tingkat akurasi sebesar 77.79%. Kemudian penelitian juga dilakukan oleh Koesriputranto pada tugas akhir yang berjudul Prediksi Harga Saham di Indonesia Dengan Menggunakan Metode Hybrid *Principal Component Analysis* dan *Support Vector Machine* (PCA-SVM). Penelitian

tersebut memperoleh hasil bahwa tingkat akurasi klasifikasi metode PCA-SVM paling baik yaitu 96.667% dibandingkan SVM dan *Neural Network*.

Penelitian ini memiliki tujuan untuk mengklasifikasikan kesejahteraan rumah tangga di Kabupaten Brebes menggunakan PCA-SVM. Serta mengetahui kinerja dari metode tersebut.

## TINJAUAN PUSTAKA

### 1. Klasifikasi

Klasifikasi adalah proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data dengan tujuan untuk memperkirakan kelas yang tidak diketahui dari suatu objek. Dalam pengklasifikasian data terdapat dua proses yang dilakukan yaitu :

#### - Proses *training*

Proses *training* menggunakan *training set* untuk membangun model atau fungsi.

#### - Proses *testing*

Untuk mengetahui keakuratan model atau fungsi yang akan dibangun pada proses *training*.

### 2. Pricipal Component Analysis (PCA)

*Principal Component Analysis* (PCA) merupakan metode yang pertama kali ditemukan oleh Karl Pearson pada tahun 1901. PCA adalah teknik statistik yang digunakan untuk menyederhanakan suatu data dengan cara mentransformasi data secara linier sehingga terbentuk sistem koordinat baru dengan varians maksimum. PCA dapat digunakan untuk mereduksi dimensi suatu data tanpa mengurangi karakteristik data tersebut secara

signifikan. Tahapan untuk melakukan PCA adalah sebagai berikut. (Abdillah dkk)

1. Menghitung nilai rata-rata dari seluruh data

$$\mu = \frac{1}{m} \sum_{i=1}^m x_i \quad (1)$$

2. Menghitung matriks kovarians C

$$C = \frac{1}{m-1} \sum_{i=1}^m (x_i - \mu)(x_i - \mu)^T \quad (2)$$

3. Menghitung nilai eigen ( $\lambda$ ) dan vector eigen ( $v$ ) dari matriks kovarians C.

$$Cv = \lambda v \quad (3)$$

4. Mengurutkan vector eigen yang berkorespondensi dengan nilai eigen dari yang terbesar hingga yang terkecil. Vector eigen digunakan untuk mentransformasi data menjadi data baru dengan dimensi data yang sama atau lebih kecil.

5. Mentransformasi data dengan mengalikan data lama dengan mengalikan matriks eigen yang telah diurutkan.

$$\hat{X} = (v^*)^T X \quad (4)$$

### 3. Support Vector Machine (SVM)

Support Vector Machine (SVM) diusulkan oleh Vanpik untuk klasifikasi dua kategori atau biner. SVM memisahkan titik-titik dari kelas yang berbeda, misalkan kelas (+1) dan (-1) dengan *hyperplane* tunggal pada ruang berdimensi banyak yang pada akhirnya partisi-partisi tersebut diselesaikan secara nonlinier. Program nonlinier yang digunakan *hyperplane* adalah dengan *quadratic programming*. *Hyperplane* yang digunakan dalam pengklasifikasian biner mempunyai *decision function* sebagai berikut.

$$f(x) = x'w - \gamma \quad (5)$$

Sebanyak  $n$  objek pada pengklasifikasian dalam ruang dimensi  $R^p$  sehingga susunan data berupa matrik A berukuran  $n \times p$  dan beranggotakan tiap titik yaitu  $y_i$  terhadap kelas (+1) atau (-

1) didefinisikan pada matriks D berukuran  $n \times n$ . Algoritma SVM standar dalam pengklasifikasian adalah sebagai berikut.

(SVM <sub>$\|\cdot\|_2$</sub> ):

$$\min_{(w, \gamma, y) \in R^{p+1+n}} v e' y + \frac{1}{2} \|w\|_2^2 \quad (6)$$

Dengan kendala  $D(Aw - e\gamma) + y \geq e$   
 $y \geq 0$

Dimana :

$v$  : Parameter yang ditentukan sebagai pengontrol (*trade off*)

$y$  : Vektor variabel *slack* berukuran  $n \times 1$  yang mengukur kesalahan klasifikasi

$e$  : Vektor kolom berukuran  $n$  dan bernilai 1.

$w$  : Vektor normal berukuran  $p \times 1$ .

$\gamma$  : Nilai bias yang menentukan lokasi relative *hyperplane* terhadap kelas asli.

Pengklasifikasian SVM nonlinier dilakukan dengan mentransformasi formulasi SVM standar sebagai berikut :

$$w = A'Du \quad (7)$$

Problem nonlinier akan terbentuk jika mendistribusikan persamaan (6) dengan (2) sehingga diperoleh sebagai berikut.

$$\min_{(w, \gamma, y) \in R^{p+1+n}} v e' y + \frac{1}{2} \|A'Du\|_2^2 \quad (8)$$

Dengan kendala :

$$D(AA'Du - e\gamma) + y \geq e$$

$$y \geq 0$$

Dengan menggantikan  $A'A$  dengan kernel nonlinier  $K(A, A')$  menghasilkan *nonlinear generalized SVM* adalah sebagai berikut :

$$\min_{(w, \gamma, y) \in R^{p+1+n}} v e' y + \frac{1}{2} u'D'K(A, A')Du \quad (9)$$

Dengan kendala

$$D(K(A, A')Du - e\gamma) + y \geq e$$

$$y \geq 0$$

#### 4. Evaluasi Kinerja Klasifikasi

Evaluasi kinerja dalam klasifikasi dapat dilihat dari tingkat akurasi, sensitivitas dan spesifitas. Akurasi klasifikasi dapat dilihat dari ketepatan klasifikasi. Akurasi klasifikasi dapat diuji dengan tingkat sensitivitas dan spesifitas. Sensitivitas adalah tingkat akurasi kelas positif sedangkan spesifitas adalah tingkat akurasi kelas negatif.

**Tabel 1. Confusion Matrix Untuk Hasil Klasifikasi Biner**

Kelas prediksi	Kelas sebenarnya	
	Positif	Negatif
Positif	Tp	Fp
Negatif	Fn	Tn

Sumber : Zhu, et al (2010)

$$\text{Akurasi klasifikasi (\%)} = \frac{tp+tn}{tp+fp+tn+fn}$$

$$\text{Sensitivity (\%)} = \frac{tp}{tp+fn}$$

$$\text{Specificity (\%)} = \frac{tn}{fp+tn}$$

#### METODE PENELITIAN

Data yang digunakan dalam penelitian ini adalah data sekunder dari Badan Pusat Statistik hasil Survei Sosial Dan Ekonomi Nasional (SUSENAS) pada wilayah Kabupaten Brebes. Digunakan 13 variabel dengan jumlah responden sebanyak 955.

Variabel yang digunakan adalah :

- Status kepemilikan rumah (X1)

- Luas lantai (X2)
- Jenis lantai terluas (X3)
- Jenis atap (X4)
- Jenis dinding (X5)
- Sumber Air Minum(X6)
- Cara memperoleh air minum (X7)
- Sumber penerangan (X8)
- Bahan bakar memasak (X9)
- Fasilitas buang air besar (X10)
- Pembuangan akhir (X11)
- Aset yang dimiliki (X12)
- Program sosial yang diikuti (X13)

Dari data yang diperoleh dilakukan pengujian menggunakan PCA SVM. Langkah penelitian yang dilakukan adalah sebagai berikut.

1. Menginputkan data penelitian
2. Melakukan pengujian klasifikasi menggunakan SVM
3. Menginterpretasi hasil klasifikasi SVM
4. Pengujian PCA-SVM dengan melakukan tahapan sebagai berikut:
  - Menentukan *eigenvalue* dan *eigenvector* pada data awal
  - Melakukan seleksi fitur
  - Melakukan pengujian SVM dengan data hasil PCA
  - Menginterpretasi hasil dari PCA+SVM
5. Mengevaluasi hasil pengujian SVM dan PCA+SVM
6. Menarik kesimpulan dari hasil pengujian SVM dan PCA+SVM

#### HASIL DAN PEMBAHASAN

##### 1. Analisis Deskriptif

**Tabel 2. Statistik Deskriptif Data Numerik**

	N	Min	Maks	Mean	St. Deviasi
Luas lantai (X2)	955	3	300	28.55	24.436

**Tabel 3. Statistik Deskriptif Data Kategorik**

	N	Modus	N for Mode	Present
Status kepemilikan rumah (X1)	955	1	805	84%
Jenis lantai terluas (X3)	955	2	603	63%

	N	Modus	N for Mode	Present
Jenis atap (X4)	955	2	927	97%
Jenis dinding (X5)	955	1	862	90%
Sumber air minum (X6)	955	3	228	24%
Cara memperoleh air minum (X7)	955	5	761	80%
Sumber penerangan (X8)	955	1	830	87%
Bahan bakar memasak (X9)	955	4	790	83%
Fasilitas BAB (X10)	955	1	720	75%
Pembuangan akhir (X11)	955	1	678	71%
Asset yang dimiliki (X12)	955	1	818	86%
Program sosial yang diikuti (X13)	955	1	788	83%

a. Status Kepemilikan Rumah/Tempat

Berdasarkan tabel 3. menunjukkan nilai modus dari status kepemilikan rumah atau tempat tinggal adalah 1. Nilai modus ini menunjukkan nilai yang sering muncul atau jumlah nilai yang terbanyak dari suatu kategorik. Nilai 1 yang sering muncul menunjukkan 84% rumah tangga memiliki status tempat tinggal milik sendiri. Status tempat tinggal milik sendiri merupakan faktor ketidakmiskinan bagi suatu rumah tangga.

b. Luas lantai

Berdasarkan tabel 4.1 menunjukkan nilai rata-rata dari luas lantai tempat tinggal tiap responden adalah 28.55 dengan standar deviasi 24.436. Rumah tangga dikategorikan miskin jika memiliki luas lantai tiap anggota rumah tangga kurang dari 8m<sup>2</sup>. Dari nilai mean sebesar 28.55 atau lebih kurang 97% menunjukkan rumah tangga di Kabupaten Brebes memiliki luas lantai lebih dari 8m<sup>2</sup> sehingga merupakan faktor ketidakmiskinan rumah tangga.

c. Jenis Lantai Terluis

Berdasarkan tabel 4.2 menunjukkan nilai modus dari jenis lantai terluis adalah 2. Nilai 2 yang sering muncul menunjukkan 63% rumah tangga menunjukkan jenis lantai terluis dengan kategori lantai keramik. Jenis lantai terluis dengan kategori lantai keramik merupakan faktor ketidakmiskinan bagi suatu rumah tangga.

d. Jenis Atap

Berdasarkan tabel 4.2 menunjukkan nilai modus dari jenis atap adalah 2. Nilai 2 yang sering muncul menunjukkan 97% rumah tangga menunjukkan jenis atap dengan kategori atap genting. Jenis atap dengan kategori atap genting merupakan faktor ketidakmiskinan bagi suatu rumah tangga.

e. Jenis Dinding

Berdasarkan tabel 4.2 menunjukkan nilai modus dari jenis dinding adalah 1. Nilai 1 yang sering muncul menunjukkan 90% rumah tangga menunjukkan jenis dinding dengan kategori tembok. Jenis dinding dengan kategori tembok merupakan faktor ketidakmiskinan bagi suatu rumah tangga.

f. Sumber Air Minum

Berdasarkan tabel 4.2 menunjukkan nilai modus dari sumber air minum adalah 3. Nilai 3 yang sering muncul menunjukkan 24% rumah tangga menunjukkan sumber air minum yang digunakan adalah leding meteran. Sumber air minum leding menunjukkan rumah tangga tidak miskin bagi suatu rumah tangga.

g. Cara Memperoleh Air Minum

Berdasarkan tabel 4.2 menunjukkan nilai modus dari cara memperoleh air minum adalah 5. Nilai 5 yang sering muncul menunjukkan 80% rumah

tangga menunjukkan cara memperoleh air minum dengan tidak membeli. Cara memperoleh air minum dengan tidak membeli menunjukkan faktor kemiskinan bagi suatu rumah tangga.

h. Sumber Penerangan

Berdasarkan tabel 4.2 menunjukkan nilai modus dari sumber penerangan adalah 1. Nilai 1 yang sering muncul menunjukkan 87% rumah tangga menunjukkan sumber penerangan listrik PLN dengan meteran. Sumber penerangan listrik PLN dengan meteran merupakan faktor ketidakmiskinan bagi suatu rumah tangga.

i. Bahan Bakar Memasak

Berdasarkan tabel 4.2 menunjukkan nilai modus dari bahan bakar memasak adalah 4. Nilai 4 yang sering muncul menunjukkan 83% rumah tangga menunjukkan bahan bakar memasak menggunakan elpiji 3kg. Bahan bakar memasak menggunakan elpiji 3kg merupakan faktor ketidakmiskinan bagi suatu rumah tangga. Faktor kemiskinan pada penggunaan bahan bakar memasak seperti kayu bakar, arang dan minyak tanah sebesar 15%.

j. Fasilitas Buang Air Besar (BAB)

Berdasarkan tabel 4.2 menunjukkan nilai modus dari fasilitas buang air besar adalah 1. Nilai 1 yang sering muncul menunjukkan 75% rumah tangga menunjukkan kepemilikan fasilitas buang air besar yang digunakan hanya untuk anggota keluarga sendiri. Kepemilikan fasilitas buang air besar sendiri merupakan faktor ketidakmiskinan bagi suatu rumah tangga.

k. Tempat Pembuangan Akhir

Berdasarkan tabel 4.2 menunjukkan nilai modus dari tempat pembuangan akhir adalah 1. Nilai 1 yang sering muncul menunjukkan 71% rumah tangga menunjukkan pembuangan akhir menggunakan tangki septik. Pembuangan akhir menggunakan tangki

septik merupakan faktor ketidakmiskinan bagi suatu rumah tangga.

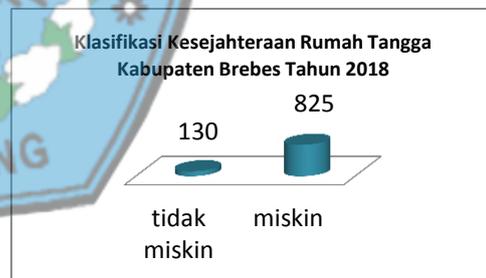
1. Asset yang Dimiliki

Berdasarkan tabel 4.2 menunjukkan nilai modus dari asset yang dimiliki adalah 1. Nilai 1 yang sering muncul menunjukkan 86% rumah tangga menunjukkan tidak memiliki asset. Ketidakpemilikan asset menunjukkan faktor kemiskinan bagi suatu rumah tangga.

m. Program Sosial yang diikuti

Berdasarkan tabel 4.2 menunjukkan nilai modus dari program sosial yang diikuti adalah 1. Nilai 1 yang sering muncul menunjukkan 83% rumah tangga mengikuti program sosial. Keikutsertaan dalam program sosial menunjukkan faktor kemiskinan bagi suatu rumah tangga.

Dari data diatas diperoleh klasifikasi kesejahteraan rumah tangga dari data awal sebagai berikut.



Gambar 1. Klasifikasi data awal

Klasifikasi awal ini dilakukan untuk mengetahui kondisi sebenarnya dari setiap rumah tangga. Terdapat 825 responden tergolong miskin dan 130 tergolong tidak miskin. Rumah tangga dengan kategori miskin terdapat sebanyak 86% dan rumah tangga tidak miskin sebanyak 14%.

2. Klasifikasi dengan SVM

Hasil klasifikasi kesejahteraan rumah tangga di Kabupaten Brebes

tahun 2018 menggunakan SVM dengan pendekatan kernel RBF sebagai berikut.

**Tabel 4. Tabel Kontingensi SVM Berdasarkan Pembagian Data**

Kelas prediksi	Pembagian data	Kelas sebenarnya			
		Training		Testing	
		Tidak miskin	miskin	Tidak miskin	Miskin
Tidak miskin	70:30	87	0	29	1
Miskin		5	576	9	248
Tidak miskin	80:20	102	1	21	0
Miskin		3	658	4	166
Tidak miskin	90:10	112	1	12	0
Miskin		4	743	2	81

Dari tabel 4.3 dapat dilakukan perhitungan akurasi, spesifitas dan sensitivitas pada data *training* menggunakan SVM dengan pembagian data 70:30 adalah sebagai berikut.

$$\text{Akurasi klasifikasi} = \frac{87+576}{87+0+5+576} \times 100\% = 99.25\%$$

$$\text{Sensitivity} = \frac{87}{87+5} \times 100\% = 94.57\%$$

$$\text{Specificity} = \frac{576}{0+576} \times 100\% = 100\%$$

Sedangkan untuk perhitungan pada data *testing* adalah sebagai berikut.

$$\text{Akurasi klasifikasi} = \frac{29+248}{29+1+9+248} \times 100\% = 96.52\%$$

$$\text{Sensitivity} = \frac{29}{29+9} \times 100\% = 76.32\%$$

$$\text{Specificity} = \frac{248}{1+248} \times 100\% = 99.6\%$$

Perhitungan tersebut juga dilakukan terhadap data dengan pembagian 80:20 dan 90:10.

### 3. Principal Component Analysis (PCA)

Proses PCA diawali dengan menormalkan data dan bebas data *missing*. Kemudian menentukan *principal component* (PC) yang akan digunakan sesuai kebutuhan. Dalam hal ini disebut dengan nilai *eigenvalue* yang menjelaskan keragaman data.

**Tabel 5. Nilai eigenvalue**

Comp	Data 70:30		Data 80:20		Data 90:10	
	Prop. Variance	Cumulative Variance	Prop. Variance	Cumulative Variance	Prop. Variance	Cumulative Variance
PC1	0.9669	0.9669	0.9655	0.9655	0.9653	0.9653
PC2	0.0114	0.9783	0.0120	0.9775	0.0119	0.9772
PC3	0.0069	0.9852	0.0073	0.9849	0.0074	0.9846
PC4	0.0053	0.9905	0.0056	0.9904	0.0057	0.9903
PC5	0.0034	0.9939	0.0034	0.9938	0.0033	0.9936
PC6	0.0017	0.9955	0.0017	0.9955	0.0018	0.9954
PC7	0.0014	0.9969	0.0014	0.9969	0.0015	0.9969
PC8	0.0011	0.9981	0.0011	0.9980	0.0011	0.9979

Comp	Data 70:30		Data 80:20		Data 90:10	
	Prop. Variance	Cumulative Variance	Prop. Variance	Cumulative Variance	Prop. Variance	Cumulative Variance
PC9	0.0011	0.9991	0.0011	0.9991	0.0011	0.9990
PC10	0.0006	0.9997	0.0006	0.9997	0.0007	0.9997
PC11	0.0001	0.9999	0.0002	0.9999	0.0002	0.9999
PC12	0.0001	1.0000	0.0001	1.0000	0.0001	1.0000

**Tabel 6. Nilai eigenvector**

	Data 70:30		Data 80:20		Data 90:10	
	PC1	PC2	PC1	PC2	PC1	PC2
Status kepemilikan rumah	-0.0022	-0.0083	-0.0010	-0.0032	-0.0010	-0.0032
Luas lantai	-0.9999	0.0065	-0.9999	0.0083	-0.9999	0.0083
Jenis lantai	0.0042	0.4413	0.0059	0.4620	0.0059	0.4620
Jenis dinding	0.0018	0.1283	0.0023	0.1201	0.0023	0.1201
Sumber air minum	0.0081	0.3830	0.0070	0.4174	0.0070	0.4174
Cara memperoleh air minum	0.0009	0.2572	0.0002	0.3141	0.0002	0.3141
Sumber penerangan	0.0011	0.0414	0.0019	0.0492	0.0019	0.0492
Bahan bakar memasak	-0.0002	0.7131	0.0012	0.6510	0.0012	0.6510
Fasilitas BAB	0.0039	0.2463	0.0046	0.2621	0.0046	0.2621
Pembuangan akhir	-0.0024	-0.0677	-0.0023	-0.0586	-0.0023	-0.0586
Asset yang dimiliki	0.0005	0.0082	0.0009	0.0135	0.0009	0.0135
Program sosial yang diikuti	-0.0031	-0.0160	-0.0036	-0.0246	-0.0036	-0.0246

Terlihat pada tabel 5 data 70:30 menunjukkan komponen pertama (PC1) yang memiliki keragaman data sebesar 96.69%. Skor untuk komponen yang dibentuk bisa dihitung dengan melihat nilai koefisien masing-masing variabel.

Komponen kedua (PC2) dapat menjelaskan 1.14% keragaman data. PC1 dan PC2 jika digabungkan akan merepresentasikan 97.83% dari keragaman total. Begitu seterusnya pada pembagian data lainnya.

**Tabel 7. Hasil Seleksi Fitur dengan PCA**

Pembagian Data	Cumulative Variance	Principal Component	Fitur Terpilih $PC \geq (\pm 0.100)$	Jumlah Fitur
70:30	97.83%	PC1, PC2	Luas lantai, jenis lantai, jenis dinding, sumber air minum, cara memperoleh air minum, bahan bakar memasak, fasilitas BAB	7
80:20	97.75%	PC1, PC2	Luas lantai, jenis lantai, jenis dinding, sumber air minum, cara memperoleh air minum, bahan bakar memasak, fasilitas BAB	7
90:10	97.72%	PC1, PC2	Luas lantai, jenis lantai, jenis dinding, sumber air minum,	7

Pembagian Data	Cumulative Variance	Principal Component	Fitur Terpilih $PC \geq (\pm 0.100)$	Jumlah Fitur
			cara memperoleh air minum, bahan bakar memasak, fasilitas BAB	

#### 4. *Principal Component Analysis – Support Vector Machine (PCA-SVM)*

Dari proses PCA yang mereduksi data kemudian dilanjutkan proses klasifikasi menggunakan SVM. Berikut hasil klasifikasi SVM dengan data hasil PCA.

**Tabel 8 Tabel Kontingensi PCA-SVM Berdasarkan Pembagian Data**

Kelas prediksi	Pembagian data	Kelas sebenarnya			
		Training		Testing	
		Tidak miskin	Miskin	Tidak miskin	miskin
<b>Tidak miskin</b>	70:30	93	1	35	0
<b>Miskin</b>		0	574	2	250
<b>Tidak miskin</b>	80:20	103	1	24	0
<b>Miskin</b>		0	660	3	164
<b>Tidak miskin</b>	90:10	118	1	10	0
<b>Miskin</b>		1	739	1	85

Dari tabel 8 dapat dilakukan perhitungan akurasi, spesifitas dan sensitivitas pada data *training* menggunakan PCA-SVM dengan pembagian data 70:30 adalah sebagai berikut.

$$\text{Akurasi klasifikasi} = \frac{93+574}{93+1+574+0} \times 100\% = 99.85\%$$

$$\text{Sensitivity} = \frac{93}{93+0} \times 100\% = 100\%$$

$$\text{Specificity} = \frac{574}{1+574} \times 100\% = 99.83\%$$

Sedangkan untuk perhitungan pada data *testing* adalah sebagai berikut.

$$\text{Akurasi klasifikasi} = \frac{35+250}{35+0+250+2} \times 100\% = 99.3\%$$

$$\text{Sensitivity} = \frac{35}{35+2} \times 100\% = 94.59\%$$

$$\text{Specificity} = \frac{250}{0+250} \times 100\% = 100\%$$

Perhitungan tersebut juga dilakukan terhadap data dengan pembagian 80:20 dan 90:10.

#### 5. **Evaluasi Kinerja SVM dan PCA-SVM**

Dari hasil klasifikasi untuk menentukan kesejahteraan rumah tangga Kabupaten Brebes tahun 2018 menggunakan metode SVM dan PCA-SVM disajikan dalam bentuk tabel 9 untuk data *training* dan data *testing*.

**Tabel 9. Evaluasi Kinerja SVM dan PCA-SVM pada Data Training**

Evaluasi Kinerja	Pembagian data	Data Training		Data Testing	
		SVM (%)	PCA-SVM (%)	SVM (%)	PCA-SVM (%)
Akurasi	70:30	99.25	<b>99.85</b>	96.52	<b>99.3</b>
	80:20	99.48	<b>99.87</b>	97.91	<b>98.43</b>
	90:10	99.42	<b>99.77</b>	97.89	<b>98.96</b>
Sensitivity	70:30	94.57	<b>100</b>	76.32	<b>94.59</b>
	80:20	97.14	<b>100</b>	84	<b>88.89</b>
	90:10	96.55	<b>99.16</b>	85.71	<b>90.91</b>
Specitivity	70:30	<b>100</b>	99.83	99.60	<b>100</b>
	80:20	99.85	99.85	100	<b>100</b>
	90:10	<b>99.87</b>	99.86	100	<b>100</b>

Keterangan : Nilai yang bercetak tebal menunjukkan nilai tertinggi pada setiap kelompok data

Secara keseluruhan metode SVM menggunakan PCA lebih baik daripada tidak menggunakan PCA.

#### KESIMPULAN

1. Kesejahteraan rumah tangga di Kabupaten Brebes tahun 2018 ditinjau melalui 13 variabel prediktor. Jika ditinjau secara personal akan menunjukkan faktor ketidakmiskinan. Secara keseluruhan diperoleh klasifikasi kesejahteraan rumah tangga sebanyak 825 rumah tangga yang dikategorikan miskin dan 130 rumah tangga yang dikategorikan tidak miskin.
2. Klasifikasi kesejahteraan rumah tangga di Kabupaten Brebes tahun 2018 menggunakan PCA-SVM secara keseluruhan lebih baik daripada menggunakan SVM.
3. Evaluasi Kinerja hasil klasifikasi dari pembagian data training dan data testing diperoleh :
  - Tingkat akurasi pada data training 80 lebih tinggi yaitu

sebesar 99.87%. Sedangkan pada data testing, data testing 30 lebih baik yaitu sebesar 99.3%.

- Tingkat sensitivitas pada data training 80 dan 70 lebih tinggi yaitu sebesar 100%. Sedangkan pada data testing, data testing 30 lebih baik yaitu sebesar 94.59%.
- Tingkat spesifitas pada data training 90 lebih baik yaitu sebesar 99.86%. Sedangkan pada data testing, data testing 30, 20 dan 10 memiliki tingkat spesifisitas sebesar 100%.
- Secara keseluruhan, kinerja pada data training terbaik pada pembagian data 80:20. Sedangkan kinerja pada data testing terbaik pada pembagian data 70:30.

## DAFTAR PUSTAKA

- Abdillah dkk., Pembelajaran Mesin Menggunakan *Principal Component Analysis* dan *Support Vector Machines* untuk Mendeteksi Diabetes. *J. Matem. Sains*, 2019,24,10-14.
- Badan Pusat Statistik. 2018. *Indikator Kesejahteraan Masyarakat*. November .BPS Republik Indonesia. Indonesia
- Cortes, C. & Vanpik, V. 1995. Support-Vector Networks. *Machine Learning*, 1 September, 20(3), pp. 273-297.
- Huang, et al. 2007. Model Selection For Support Vector Machines Via Uniform Design. *Elsevier* (12) : 335-346.
- Koesriputranto, Astandri. 2015. Prediksi Harga Saham di Indonesia Dengan Menggunakan Metode Hybrid *Principal Component Analysis* dan *Support Vector Machine* (PCA-SVM). *Skripsi*. Jurusan Sistem Informasi Institut Teknologi Sepuluh Nopember. Surabaya.
- Mangasarian, O.L. 1998. Generalized Support Vector Machines. *Researchgate* (13) : 1-13.
- +
- Sahitayakti, R., dan Fithriasari, K. 2015. Klasifikasi Kesejahteraan Rumah Tangga di Provinsi Papua dengan Metode Regresi Logistik dan Support Vector Machine. *J. Sains Dan Seni ITS* 4(2) : 2-8.
- Vatih, Ibrahim.2009.Peta Kemiskinan. Retrieved Juli 29<sup>th</sup>, 2010, from <http://www.imz.or.id/new/article/335/peta-kemiskinan/?lang=id>
- Zhu, Wen., et al. 2010. Sensitivity, Specificity, Accuracy, Associated Confidence Interval and ROC Analysis with Practical SAS®Implementations. *Health Care and Life Sciences*(9) : 1-9.