JUNIF: Jurnal Nasional Informatika

OPEN ACCES p-ISSN: 2774-3497 e-ISSN: 2774-3500

Vol. 2, No. 1, April 2021, pp. 29-34

PREDIKSI PENYAKIT JANTUNG MENGGUNAKAN *SUPPORT VECTOR MACHINE* DAN *PYTHON* PADA BASIS DATA PASIEN DI CLEVELAND

PENULIS [1)]**Dwi Sidik Permana,** [2)]**Astried Silvanie**

ABSTRAK

Support Vector Machine (SVM) digunakan dalam penelitian ini untuk memprediksi penyakit jantung berdasarkan 13 kondisi medis pasien. Kondisi medis ini digunakan sebagai atribut predikator dalam penelitian ini. Keluaran yang ingin diprediksi berupa kelas target bernilai 1 jika pasien penyakit jantung dan 0 jika pasien bukan penyakit jantung. Pelatihan dilakukan dengan *Python* dan pustaka scikit. SVM diuji menggunakan empat macam kernel yaitu linear, RBF, *polynomial* dan *sigmod*. Dari hasil pelatihan model dengan nilai metrik terbaik didapatkan jika menggunakan kernel linear. Nilai metrik akurasi sama dengan 90.11%, presisi 90.38% dan *recall* 92.15% dengan kernel linear.

Kata Kunci: Prediksi Penyakit Jantung, *Machine Learning*, *Support Vector Machine*, SVM, *Python Machine Learning*

AFILIASI

Prodi, Fakultas: Teknik Informatika, Fakultas Ilmu Komputer
Nama Institusi: Institut Bisnis dan Informatika (IBI) Kosgoro 1957
Alamat Institusi: Jl. M. Kahfi II No. 33, Jagakarsa, Jakarta Selatan, DKI Jakarta

KORESPONDENSI

Penulis: Astried Silvanie
Email: astried@ibi-k57.ac.id

diterbitkan oleh:
Program Studi Teknik Informatika
Institut Bisnis dan Informatika (IBI) Kosgoro 1957

I. PENDAHULUAN

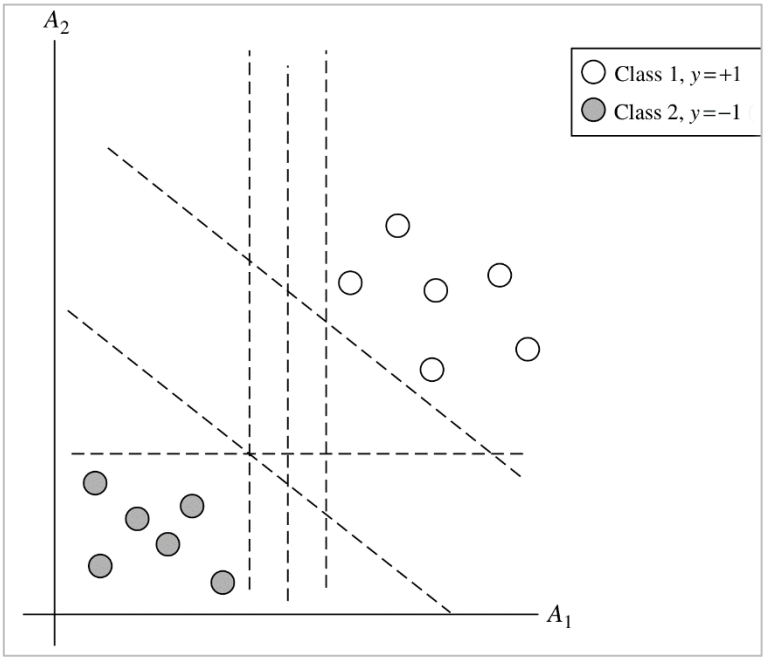
1.1 Latar Belakang

Menurut data WHO (2011), penyakit kardiovaskular adalah penyebab utama kematian secara global, mengambil sekitar 17,9 juta jiwa setiap tahun. Kardiovaskular adalah sekelompok gangguan jantung dan pembuluh darah dan termasuk penyakit jantung koroner, penyakit serebrovaskular, penyakit jantung rematik dan kondisi lainnya.

Menurut VV. Ramalingam (2018), algoritma dan teknik pembelajaran mesin telah diterapkan ke berbagai kumpulan data medis untuk mengotomatisasi analisis data yang besar dan kompleks. Teknik pembelajaran mesin banyak digunakan dalam penelitian untuk membantu industri kesehatan dalam diagnosis penyakit salah satunya penyakit jantung. Pahwa dkk (2017) menggunakan Naive Bayes untuk prediksi penyakit jantung. Pouriyeh dkk (2017) menggunakan *K–Nearest Neighbour* untuk prediksi penyakit jantung. Sedangkan dalam penelitian kali ini, kami menggunakan *Support Vector Machine* untuk memprediksi penyakit jantung berdasarkan kondisi medis pasien.

1.2 *Support Vector Machine* (SVM)

Menurut Jiawei (2006), *Support Vector Machine* adalah metode klasifikasi untuk data linear dan non linear dan termasuk ke kategori pembelajaran mesin dengan pengawasan. Metode ini menggunakan pemetaan *nonlinier* untuk mengubah data pelatihan asli ke dimensi yang lebih tinggi. Dalam dimensi baru ini akan dicari optimasi linier yang memisahkan dua kelas target dengan *hyperplane*. *Hyperplane* adalah batas keputusan yang memisahkan tipe dari satu kelas dengan kelas yang lain. SVM menemukan *hyperplane* menggunakan vektor pendukung dan margin.



Gambar 1. Hyperplane di SVM yang Terbagi Atas Dua Kelas Target

II. METODE PENELITIAN

Berikut ini langkah-langkah yang dilakukan dalam penelitian:

1) Identifikasi Masalah. Identifikasi masalah yang ingin dicari solusinya, dalam penelitian ini adalah bagaimana kita memprediksi apakah seorang pasien terkena penyakit jantung atau tidak berdasarkan kondisi medis mereka.
2) Pengumpulan dan PraProses Data. Tahap ini menentukan sumber data yang berupa data sekunder diperoleh dari Gudang Pembelajaran Mesin UCI. Data ini di kumpulan oleh Robert Detrano, M.D., Ph.D dari V.A. Medical Center, Long Beach and Cleveland Clinic Foundation.
3) Desain dan implementasi. Proses perancangan, pelatihan dan pengujian data diimplementasikan dengan bahasa *python* dan pustaka scikit.
4) Hasil: Hasil yang didapatkan adalah koefisien bobot untuk setiap atribut predikator.

III. HASIL DAN PEMBAHASAN

3.1 Data Olah

Database yang digunakan adalah basis data pasien dan kondisi klinisnya dari V.A. Medical Center, Long Beach and Cleveland Clinic Foundation. Gejala Penyakit kardiovaskular terdiri dari berbagai kondisi yang mempengaruhi kerja jantung dan vena darah dan cara darah dipompa dan diedarkan melalui tubuh (Grace, S.L skk, 2004). Basis data ini mempunyai 303 pasien dengan 14 atribut. Sebanyak 13 atribut merupakan kondisi klinis pasien yang digunakan sebagai atribut prediksi. Sedangkan kelas target adalah atribut ke-14,atribut ini mempunyai dua nilai saja. Nilai 1 untuk ya pasien terkena penyakit jantung dan 0 untuk pasien bukan penyakit jantung. Berikut ini rincian atribut dalam bentuk tabulasi.

Tabel 1. Atribut Prediksi dan Kelas Target pada Basis Data Pasien Jantung di Cleveland.

NO	ATRIBUT	TIPE DATA
x1	age	integer
x2	sex	integer
x3	chest pain type (4 values)	integer
x4	resting blood pressureinteger	integer
x5	serum cholestoral in mg/dl	integer
x6	fasting blood sugar > 120 mg/dl	integer
x7	resting electrocardiographic results (values 0,1,2)	integer
x8	maximum heart rate achieved	integer
x9	exercise induced angina	integer
x10	oldpeak = ST depression induced by exercise relative to rest	float
x11	the slope of the peak exercise ST segment	integer
x12	number of major vessels (0-3) colored by flourosopy	integer
x13	thal: 3 = normal; 6 = fixed defect; 7 = reversable defect	integer
y	Target: 1 = Penyakit Jantung, 0 = Bukan penyakit jantung	integer

3.2 Pembahasan dan Implementasi

Pertama, impor pustaka yang diperlukan yaitu pandas. Pustaka ini akan meload data yang berbentuk format comma separated values (.csv) pengubah dataset.

```
1  import pandas as pd
2
3  dataset = pd.read_csv('heart.csv',delimiter=",")
```

Pilih atribut indeks ke-0 sampai dengan indeks ke-12 sebagai atribut predikator dan simpan dalam pengubah Fitur.

```
4  Fitur = dataset.iloc[:,0:13]
5  Target = dataset.iloc[:,13]
```

Pembagian data menjadi data latihan dan data uji coba. Disini kita akan mengambil 70% untuk menjadi data latihan dan sisa 30% menjadi data uji coba.

```
6  from sklearn.model_selection import train_test_split
7  from sklearn import metrics
8
9  X_train, X_test, y_train, y_test = train_test_split(Fitur, Target, test_size=0.3,random_state=109)
10
```

Kemudian kita akan memasukkan data ke dalam fungsi SVM untuk menjalankan algoritma svm.SCV. Algoritma SVM menggunakan seperangkat fungsi matematika yang didefinisikan sebagai kernel. Fungsi kernel adalah mengambil data sebagai input dan mengubahnya menjadi ruang dimensi tinggi disebut *kernel*

space. Algoritma SVM yang berbeda menggunakan berbagai jenis fungsi kernel seperti linear, nonlinier, polinomial, *radial basis function* (RBF), dan sigmoid (Awwad dkk, 2017). Dalam penelitian ini kita menguji SVM dengan empat macam kernel yaitu:

1) Linear
   $K(x, u) = x^T . u$

2) Polynomial
   $K(x, u) = (ax^T u + c)^q, q > 0$

3) Gaussian radial basis function (RBF)
   $$K(x,u) = \exp\left(-\frac{\|x-u\|^2}{\sigma^2}\right)$$

4) Sigmod
   $K(x, u) = \tanh(ß\, x^T u + ∂)$
   (Awwad dkk, 2017)

```
#Kernel Linear
from sklearn import svm
from sklearn import metrics

clf = svm.SVC(kernel='linear')
clf.fit(X_train, y_train)
y_pred = clf.predict(X_test)

```

```
#RBF Linear
from sklearn import svm
from sklearn import metrics

clf = svm.SVC(kernel='rbf')
clf.fit(X_train, y_train)
y_pred = clf.predict(X_test)

```

```
#Poly Linear
from sklearn import svm
from sklearn import metrics

clf = svm.SVC(kernel='poly')
clf.fit(X_train, y_train)
y_pred = clf.predict(X_test)
```

```
#Kernel Sigmod
from sklearn import svm
from sklearn import metrics

clf = svm.SVC(kernel='sigmoid')
clf.fit(X_train, y_train)
y_pred = clf.predict(X_test)
```

IV. PENUTUP

Kita mengukur kinerja model dengan tiga metrik, yaitu akurasi, *recall* dan *precision*. Akurasi dalam klasifikasi adalah rasio jumlah prediksi yang benar dari semua prediksi yang dibuat. *Recall* atau Sensitivitas adalah proporsi kasus *True Positive* (TP) yang diprediksi Positif dengan memang benar. Ukuran ini mengukur cakupan kasus TP dengan aturan +P (diprediksi positif). Fitur yang diinginkan adalah mencerminkan berapa banyak kasus relevan yang diambil oleh aturan +P benar (Powers dkk, 2011). *Precision* atau *Confidence* menunjukkan proporsi kasus Positif Terprediksi yang benar *True Positive* (TP) (Powers dkk, 2011).

True Positive (TP), adalah jumlah pasien yang diklasifikasikan memiliki penyakit jantung dan memang benar memiliki penyakit jantung.

True Negative (TN), adalah jumlah pasien yang diklasifikasikan sebagai tidak berpenyakit jantung dan memang tidak memiliki penyakit jantung.

False Negative (FN), adalah jumlah pasien yang diklasifikasikan sebagai tidak memiliki penyakit jantung padahal pasien tersebut memang memiliki penyakit jantung.

False Positive (FP), adalah jumlah pasien yang diklasifikasikan memiliki penyakit jantung padahal mereka sebenarnya tidak memiliki penyakit jantung.

Perhitungan akurasi, presisi dan specify adalah sebagai berikut:

Accuracy = (TP+TN) / (TP+TN+FP+FN)
Recall = TP / (TP+FN)
Precision = TP / (TP+FP)

Tabel 2. Metrik Pengukuran Kinerja Model SVM

KERNEL	ACCURACY	RECALL	PRECISION
Linear	90.11%	92.15%	90.38%
Polynomial	67.03%	76.47%	68.42%
Radial Basis Function (RBF)	69.23%	86.27%	67.69%
Sigmoid	56.04%	1%	56.04%

Dari tabel kita dapat melihat kernel linear memberikan hasil terbaik dan ini menunjukkan jika data kita berbentuk linear. Nilai koefisien yang didapatkan dari kernel linear dapat dilihat pada tabel di bawah.

Tabel 3. Hasil Koefisien Bobot Untuk Setiap Atribut Predikator

KONDISI KLINIS	KOEFISIEN BOBOT
age	-0.01407653
sex	-0.868647
chest pain	0.56661076
resting blood pressure	-0.01639903
serum cholestoral in mg/dl	-0.00139516
fasting blood sugar > 120 mg/dl	0.28524992
resting electrocardiographic results (values 0,1,2)	0.27133202
maximum heart rate achieved	0.01342528
exercise induced angina	-0.42898128
oldpeak = ST depression induced by exercise relative to rest	-0.27638487
the slope of the peak exercise ST segment	0.24566734
number of major vessels (0-3) colored by flourosopy	-0.52802942
thal: 3 = normal; 6 = fixed defect; 7 = reversable defect	-0.76581845

DAFTAR PUSTAKA

Awad, Mariette & Khanna, Rahul. (2015). Support Vector Machines for Classification. 10.1007/978-1-4302-5990-9_3.

Chauhan, Raj H., Daksh N. Naik., Rinal A. Halpati., Sagarkumar J. Patel. & Mr. A.D.Prajapati. (2020). Disease Prediction using Machine Learning. International Research Journal of Engineering and Technology (IRJET) Volume: 07 Issue: 05 | May 2020.

Farooqui, Md. Ehtisham and Ahmad, Dr. Jameel, Disease Prediction System using Support Vector Machine and Multilinear Regression (August 13, 2020). International Journal of Innovative Research in Computer Science & Technology (IJIRCST) ISSN: 2347-5552, Volume, 8, Issue, 4, July, 2020.

Grace, S.L., Rick Fry , Angela Cheung & Donna E Stewart. (2004). Cardiovascular Disease. *BMC Women's Health* 4**,**S15. https://doi.org/10.1186/1472-6874-4-S1-S15.

Han, Jiawei. dan Michael Kamber. (2006). Data Mining Concept and Techniques, 2nd edition, USA: Elsevier, Inc, 2006.

Kramar, Vadym & Alchakov, Vasiliy & Dushko, Veronika & Kramar, Tatiana. (2018). Application of support vector machine for prediction and classification. Journal of Physics: Conference Series. 1015. 032070. 10.1088/1742-6596/1015/3/032070.

Pahwa, Kanika & Ravinder Kumar dkk. (2017). Prediction of Heart Disease Using Hybrid Technique For Selecting Features, 2017 4th IEEE Uttar Pradesh Section International Conference on Electrical, Computer and Electronics (UPCON).

Pouriyeh, Seyedamin., Sara Vahid., Giovanna Sannino., Giuseppe De Pietro., Hamid Arabnia., Juan Gutierrez. (2017). A Comprehensive Investigation and Comparison of Machine Learning Techniques in the Domain of Heart Disease. 22nd IEEE Symposium on Computers and Communication (ISCC 2017): Workshops - ICTS4eHealth 2017.

Powers, David & Ailab,. (2011). Evaluation: From precision, recall and F-measure to ROC, informedness, markedness & correlation. J. Mach. Learn. Technol. 2. 2229-3981. 10.9735/2229-3981.

Rufai, Ahmad., U. S. Idriss & Mahmood Umar. (2018). Using Artificial Neural Networks to Diagnose Heart Disease. International Journal of Computer Applications. 182. 1-6. 10.5120/ijca2018917938.

Thirugnanam, Mythili. (2013). A Heart Disease Prediction Model using SVM-Decision Trees-Logistic Regression (SDL). International Journal of Computer Applications in Technology. 68. 11-15. 10.5120/11662-7250.

V.V. Ramalingam, Ayantan Dandapath, M Karthik Raja. (2018). Heart disease prediction using machine learning techniques: a survey. International Journal of Engineering & Technology, 7 (2.8) (2018) 684-687

WHO. (2011). *Global Atlas on Cardiovascular Disease Prevention and Control*. World Health Organization; 2011.

Yihua, Zhong., Zhao Lei, Liu Zhibin, Xu Yao & Li Rong. (2010). Using a support vector machine method to predict the development indices of very high water cut oilfi elds. Petroleum Science, 2010 – Springer.