

**PREDIKSI PERSEBARAN DEMAM BERDARAH DI KOTA BANDAR  
LAMPUNG DENGAN METODE *SUPPORT VECTOR MACHINE***

(SKRIPSI)

Oleh:

RM Sulaiman Sani



**S1 ILMU KOMPUTER  
JURUSAN ILMU KOMPUTER  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS LAMPUNG  
2019**

## **ABSTRACT**

### **PREDICTION OF BLOOD FEVER IN BANDAR LAMPUNG CITY USING VECTOR MACHINE SUPPORT METHOD**

**By**

**RM SULAIMAN SANI**

Dengue fever is a dangerous disease and always causes casualties every year, especially in the tropics or subtropics. Dengue fever cases increase during the rainy season, many factors affect the spread of dengue fever, such as vegetation, population and landfills. The data used are dengue data in Bandar Lampung City, weather data, population data and distance matrix data between dengue fever events with each other. The amount of data used is 1,080 data with 3 kernels: linear, gaussian and polynomial. In this study four experiments were carried out, the first two experiments were carried out without Feature Selection and the next two experiments were carried out with Feature Selection. After the experiment was found the highest accuracy in the experiment with Feature Selection with 44 Variables. From the experiments conducted on the Gaussian kernel get the highest accuracy that is equal to 75.52%, while the Linear kernel accuracy gets the lowest accuracy that is equal to 74.61% and on the Polynomial Kernel get an accuracy of 75.15%.

**Keywords :** Machine Learning, Prediction of Dengue Fever, Support Vector Machine.

## **ABSTRAK**

### **PREDIKSI PERSEBARAN DEMAM BERDARAH DI KOTA BANDAR LAMPUNG DENGAN METODE *SUPPORT VECTOR MACHINE***

**Oleh**

**RM SULAIMAN SANI**

Demam berdarah adalah penyakit berbahaya dan selalu menyebabkan korban setiap tahun, terutama di daerah tropis atau subtropis. Kasus demam berdarah meningkat selama musim hujan, banyak faktor yang mempengaruhi penyebaran penyakit demam berdarah, seperti vegetasi, jumlah penduduk dan tempat pembuangan sampah. Data yang digunakan adalah data penderita demam berdarah di Kota Bandar Lampung, data cuaca, data penduduk dan data *matriks* jarak antara kejadian demam berdarah satu sama lain. Jumlah data yang digunakan adalah 1.080 data dengan 3 kernel: linear, gaussian dan polinomial. Dalam penelitian ini dilakukan empat kali eksperimen, dua eksperimen pertama dilakukan tanpa *Feature Selection* dan dua eksperimen berikutnya dilakukan dengan *Feature Selection*. Setelah dilakukan percobaan didapati akurasi tertinggi pada percobaan dengan Feature Selection dengan 44 *variable*. Dari percobaan yang dilakukan pada kernel Gaussian mendapatkan akurasi tertinggi yaitu sebesar 75,52%, sedangkan akurasi pada kernel Linear mendapatkan akurasi terendah yaitu sebesar 74,61% dan pada Kernel Polynomial mendapatkan akurasi sebesar 75,15%.

**Kata Kunci** : *Machine Learning*, Prediksi demam berdarah, *Support Vector Machine*.

**PREDIKSI PERSEBARAN DEMAM BERDARAH DI KOTA BANDAR  
LAMPUNG DENGAN METODE *SUPPORT VECTOR MACHINE***

**Oleh  
RM SULAIMAN SANI**

**Skripsi  
Sebagai Salah Satu Syarat untuk Memperoleh Gelar  
SARJANA KOMPUTER**

**Pada**

**Jurusan Ilmu Komputer  
Fakultas Matematika dan Ilmu Pengetahuan Alam**



**S1 ILMU KOMPUTER  
JURUSAN ILMU KOMPUTER  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS LAMPUNG  
2019**

Judul Skripsi

: **PREDIKSI PERSEBARAN DEMAM  
BEDARAH DI KOTA BANDAR LAMPUNG  
DENGAN METODE *SUPPORT VECTOR  
MACHINE***

Nama Mahasiswa

: **RM Sulaiman Sani**

Nomor Pokok Mahasiswa

: 1517051180

Jurusan

: Ilmu Komputer

Fakultas

: Matematika dan Ilmu Pengetahuan Alam



**Favorisen R. Lumbanraja, Ph.D**  
NIP. 19830110 200812 1 002

2. Ketua Jurusan Ilmu Komputer

**Dr. Ir. Kurnia Muludi, M.S.Sc**  
NIP. 19640616 198902 1001



## MENGESAHKAN

### 1. Tim Penguji

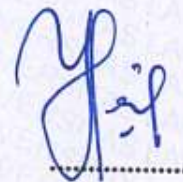
Ketua Penguji : Favorisen R. Lumbanraja, Ph.D.



Penguji Pembahas : Didik Kurniawan, M.T.



Penguji Pembahas : Anie Rose Irawati, ST, M.Cs.



2. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam



  
Drs. Suratman, M.Sc.  
NIP. 19640604 199003 1 002

Tanggal Lulus Ujian Skripsi : 12 Agustus 2019

## PERNYATAAN

Saya yang bertanda tangan di bawah ini, menyatakan bahwa skripsi saya yang berjudul “Prediksi Persebaran Demam Berdarah Di Kota Bandar Lampung Dengan Metode *Support Vector Machine*” merupakan karya saya sendiri dan bukan karya orang lain. Semua tulisan yang tertuang di skripsi ini telah mengikuti kaidah penulisan karya ilmiah Universitas Lampung. Apabila dikemudian hari terbukti skripsi saya merupakan hasil penjiplakan atau dibuat orang lain, maka saya bersedia menerima sanksi berupa pencabutan gelar yang telah saya terima.

Bandar Lampung, 12 Agustus 2019



RM SULAIMAN SANI  
NPM. 1517051180

## **RIWAYAT HIDUP**



Penulis lahir di Bandar Lampung, 17 Oktober 1997, sebagai anak pertama dari tiga bersaudara, putra dari Bapak R M Syafei dan ibu Neni Triana.

Jenjang pendidikan formal dimulai di TK Al-Anwar, Sukabumi, Bandar Lampung yang diselesaikan pada tahun 2003. Selanjutnya penulis melanjutkan pendidikan di SD Negeri 1 Campang Raya, Sukabumi, Bandar Lampung dan diselesaikan pada tahun 2009, kemudian penulis melanjutkan pendidikan di SMP Negeri 5 Bandar Lampung dan diselesaikan pada tahun 2012. Selanjutnya melanjutkan pendidikan di SMA Negeri 5 Bandar Lampung yang diselesaikan pada tahun 2015. Pada tahun yang sama, penulis diterima sebagai mahasiswa, Program Studi Ilmu Komputer, Fakultas Matematika Dan Ilmu Pengetahuan Alam Universitas Lampung melalui jalur Mandiri.

Pada tahun 2017, penulis melaksanakan Kuliah Kerja Lapangan (KKL) berupa kunjungan pendidikan ke Bandung dan Jakarta. Pada tahun 2018, penulis melaksanakan program Kuliah Kerja Nyata (KKN) di Desa Kebon Damar, Kabupaten Lampung.



## **MOTTO**

“Karena sesungguhnya sesudah kesulitan itu ada kemudahan.

Sesungguhnya sesudah kesulitan itu ada kemudahan”

[Q.S. Al-Insyirah: 5-6]

“Jika kau menungguku untuk menyerah, kau akan

menungguku selamanya”

[Naruto Uzumaki]

“Lakuin aja dulu”

[RM Sulaiman Sani]

## **PERSEMBAHAN**

Puji syukur kehadiran Allah SWT yang selalu memberikan limpahan rahmat dan karunia-Nya. Dengan kerendahan hati, kupersembahkan karya yang sederhana ini sebagai tanda bakti dan kasih cintaku yang tulus dan mendalam kepada:

Kedua orang tuaku, Syafei dan Neni yang telah mendidikku, menyayangiku dan tak pernah henti untuk selalu mendo'akanku serta memberikan semangat demi keberhasilanku.

Adik tersayang, R.A Syafriani dan R.A Kesuma Cahyani yang selalu memberikan dukungan, semangat, dan juga do'a untuk keberhasilanku.

Para pendidik yang kuhormati dan Almamaterku tercinta, Universitas Lampung.

## SANWACANA

Puji dan syukur kehadiran Allah SWT, karena atas rahmat dan ridho-Nya, penulis dapat menyelesaikan skripsi ini. Skripsi dengan judul “Prediksi persebaran Demam Berdarah Di Kota Bandar Lampung Dengan Metode *support Vector Machine*” adalah salah satu syarat untuk memperoleh gelar Sarjana Ilmu Komputer di Universitas Lampung.

Dalam kesempatan ini penulis mengucapkan terima kasih kepada:

1. Bapak Favorisen R. Lumbanraja S.Kom., M.Si., Ph.D. sebagai Pembimbing utama yang telah memberikan motivasi, pembelajaran, ilmu, semangat serta bimbingan sehingga penulisan skripsi ini dapat diselesaikan.
2. Bapak Didik Kurniawan, M.T. sebagai penguji I skripsi. Terima kasih atas masukan, saran, serta bimbingan dalam penyelesaian skripsi ini.
3. Ibu Anie Rose Irawati, ST, M.Cs selaku Penguji II skripsi. Terima kasih atas masukan, saran, serta bimbingan dalam penyelesaian skripsi ini.
4. Bapak Drs. Suratman, M.Sc selaku Dekan FMIPA Universitas Lampung.
5. Bapak Dr. Ir. Kurnia Muludi, M.S.Sc. selaku Ketua Jurusan Ilmu Komputer
6. Teman bahagia dan mitra motivasiku Dwi Kusdayanti yang senantiasa menghibur dan menemani disela-sela waktu selama mengerjakan skripsi.
7. Sahabat perkuliahanku Tawon Kunir yang telah kebersamai dari awal perkuliahan hingga sekarang.

Penulis berdoa'a semoga semua kebaikan yang telah kalian bagi untuk penulis,  
Allah gantikan dengan pahala, dan semoga skripsi ini dapat bermanfaat bagi  
semua yang membacanya. Aamiin ya rabbal 'alamiin.

## DAFTAR ISI

|  | Halaman |
|--|---------|
| DAFTAR ISI .....                                       | xiv     |
| DAFTAR TABEL .....                                     | xiv     |
| DAFTAR GAMBAR .....                                    | xviii   |
| I. PENDAHULUAN   |         |
| A. Latar Belakang Masalah .....                        | 1       |
| B. Rumusan Masalah .....                               | 4       |
| C. Tujuan .....  | 4       |
| D. Batasan Masalah .....                               | 4       |
| E. Manfaat .....                                       | 5       |
| II. TINJAUAN PUSTAKA                                   |         |
| A. <i>Machine Learning</i> .....                       | 6       |
| B. Alur Kerja Machine Learning .....                   | 7       |
| C. <i>Machine Learning</i> Berdasarkan Tipe Data ..... | 7       |
| 1. <i>Supervised Learning</i> .....                    | 7       |
| a. <i>Reggesion</i> .....                              | 8       |
| b. <i>Classification</i> .....                         | 8       |
| 2. <i>Unsupervised Learning</i> .....                  | 8       |
| D. <i>Cross validation</i> .....                       | 8       |
| 1. <i>k-fold crossvalidation</i> .....                 | 8       |
| 2. Stratified K-Fold Cross Validation .....            | 9       |
| 3. Leave-P-Out Cross Validation .....                  | 9       |



|   |    |
|---|----|
| E. SVM( <i>Support vector Machine</i> ..... | 10 |
| 1. Kernel SVM Linear .....                  | 15 |
| 2. Kernel SVM Gaussian .....                | 16 |
| 3. Kernel SVM Polynomial .....              | 17 |
| F. Uji Koefisien Determinasi .....          | 18 |
| 1. Sum Of Squares Error .....               | 18 |
| 2. Total sum of square .....                | 18 |
| 3. R-Squared .....                          | 19 |

### III. METODE PENELITIAN

|                                      |    |
|--------------------------------------|----|
| A. Waktu dan Tempat Penelitian ..... | 20 |
| 1. Tempat Penelitian .....           | 20 |
| 2. Waktu penelitian .....            | 20 |
| B. Data dan Alat .....               | 21 |
| 1. Data .....                        | 21 |
| 2. Alat .....                        | 26 |
| C. Metode implementasi .....         | 26 |
| 1. Pengumpulan Data .....            | 27 |
| 2. Pemodelan dengan SVM .....        | 28 |
| 3. Pengujian Model .....             | 28 |

### IV. HASIL DAN PEMBAHASAN

|   |    |
|---|----|
| A. Percobaan Tanpa <i>Feature Selection</i> ..... | 30 |
| 1. Percobaan Tanpa Matrix Jarak.....              | 30 |
| a. Import Data .....                              | 30 |
| b. Pemrosesan Data .....                          | 31 |
| c. Pembagian Data Set .....                       | 32 |
| d. Membuat Model Regresi .....                    | 33 |
| e. Prediksi Dengan SVM .....                      | 33 |
| 2. Percobaan Dengan Matrix Jarak.....             | 36 |
| a. Import Data .....                              | 36 |
| b. Pemrosesan Data .....                          | 37 |

|   |    |
|---|----|
| c. Pembagian Data Set .....                       | 38 |
| d. Membuat Model Regresi .....                    | 39 |
| e. Prediksi Dengan SVM .....                      | 39 |
| B. Percobaan Tanpa <i>Feature Selection</i> ..... | 42 |
| 1. Percobaan Tanpa Matrix Jarak.....              | 42 |
| a. Import Data .....                              | 42 |
| b. Pemrosesan Data .....                          | 43 |
| c. Pembagian Data Set .....                       | 44 |
| d. Membuat Model Regresi .....                    | 44 |
| e. Prediksi Dengan SVM .....                      | 45 |
| 2. Percobaan Dengan Matrix Jarak.....             | 47 |
| a. Import Data .....                              | 47 |
| b. Pemrosesan Data .....                          | 49 |
| c. Pembagian Data Set .....                       | 49 |
| d. Membuat Model Regresi .....                    | 50 |
| e. Prediksi Dengan SVM .....                      | 51 |
| C. Pembahasan .....                               | 53 |
| D. Pebandingan Dengan Penelitian Terdahulu .....  | 58 |
| <br>V. SIMPULAN DAN SARAN                         |    |
| A. Simpulan .....                                 | 60 |
| B. Saran .....                                    | 61 |
| <br>DAFTAR PUSTAKA .....                          | 52 |

## DAFTAR TABEL

| <b>Tabel</b>  | <b>Halaman</b> |
|---|----------------|
| 1. Contoh Data DBD Puskesmas Bakung .....                                   | 16             |
| 2. Contoh Data yang Di Gunakan Dalam<br>Percobaan Tanpa mtrix Jarak .....   | 24             |
| 3. Data yang sudah diubah nama atributnya .....                             | 25             |
| 4. Hasil Percobaan .....  | 27             |
| 5. Contoh Data yang Di Gunakan Dalam<br>Percobaan Dengan mtrix Jarak .....  | 30             |
| 6. Data yang sudah diubah nama atributnya .....                             | 31             |
| 7. Hasil Percobaan .....  | 33             |
| 8. Contoh Data yang Di Gunakan Dalam<br>Percobaan Tanpa mtrix Jarak .....   | 35             |
| 9. Data yang sudah diubah nama atributnya .....                             | 36             |
| 10. Hasil Percobaan .....   | 39             |
| 11. Contoh Data yang Di Gunakan Dalam<br>Percobaan Dengan mtrix Jarak ..... | 41             |
| 12. Data yang sudah diubah nama atributnya .....                            | 42             |
| 13. Hasil Percobaan .....   | 45             |

## DAFTAR GAMBAR

| Gambar   | Halaman |
|--|---------|
| 1. Pembagian data menjadi 2 <i>class</i> .....   | 9       |
| 2. Proses prediksi dengan model SVM .....  | 20      |
| 3. Cara kerja <i>k-fold cross validation</i> .....   | 22      |
| 4. Perbandingan data aktual dengan data hasil prediksi dengan<br>Metode SVM Linear .....     | 49      |
| 5. Perbandingan data aktual dengan data hasil prediksi dengan<br>Metode SVM Gaussian.....    | 49      |
| 6. Perbandingan data aktual dengan data hasil prediksi dengan<br>Metode SVM Polynomial ..... | 50      |
| 7. Perbandingan Hasil Akurasi dengan Penelitian Terdahulu .....                              | 52      |

## **I. PENDAHULUAN**

### **A. Latar Belakang**

Demam berdarah merupakan suatu penyakit yang berbahaya dan selalu memakan korban setiap tahunnya terutama di daerah tropis ataupun subtropis. Peningkatan penyakit demam berdarah terjadi selama musim hujan. Namun, banyak faktor-faktor daerah yang mempengaruhi penyebaran penyakit DBD, termasuk faktor-faktor seperti vegetasi, tipe perumahan, kepadatan penduduk, fasilitas pembuangan sampah, dan sebagainya (Fanany, 2012).

Data dari Dinas Kesehatan Kota Bandar Lampung menyebutkan pada 2014, jumlah pasien demam berdarah di Bandar Lampung mencapai 763 orang dan 16 orang meninggal. Pada 2015, jumlah demam berdarah, pasien demam di Bandar Lampung mencapai 413 orang dan 7 orang meninggal. Pada tahun 2016, peningkatan jumlah pasien demam berdarah di Bandar Lampung mencapai 1.111 orang dan 11 orang yang meninggal, jumlah tersebut merupakan yang tertinggi dibandingkan dengan kabupaten lain (Fauzi et al., 2018)



*Forecasting* (peramalan) merupakan cara untuk meramalkan kejadian masa depan berdasarkan data masa lalu yang diukur secara *periodic* dan akan membentuk suatu rangkai waktu data atau yang biasa disebut dengan *Time Series*. Rangkai waktu merupakan suatu pengamatan terhadap variabel tunggal yang diukur secara teratur selama periode waktu tertentu. Alasan mengapa melakukan analisis rangkai waktu adalah untuk meramalkan atau memprediksi nilai masa depan. *Forecasting* berdasarkan jangka waktunya dapat dibagi menjadi tiga, yaitu : jangka pendek (*Short Term*) merupakan jangka pendek meliputi kurun waktu mulai dari satu hari sampai satu musim atau dapat sampai satu tahun, jangka menengah (*Medium Term*) jangka menengah meliputi kurun waktu dari satu musim sampai dua tahun dan jangka panjang (*Long Term*) jangka panjang meliputi peramalan untuk kurun waktu minimal lima tahun (Setyawan, Subantoro, & Prabowo, 2016).

Dengan mengetahui model rangkai waktu masa lalu dapat meramalkan nilai masa depan, apakah akan naik atau turun (Putro, Furqon, & Wijoyo, 2018). Persiapan untuk menghadapi gejala penyakit pada masa yang akan datang sangat penting, karena berkaitan dengan cepat atau lambatnya penanganan terhadap masyarakat itu sendiri. Jika tidak ada persiapan yang matang, maka penanganan penyakit dimasa depan akan sulit dilakukan, dan akan menyebabkan buruknya pelayanan terhadap masyarakat itu sendiri

Salah satu metode yang dapat digunakan dalam proses prediksi adalah *Support Vector Machine*. Metode ini merupakan teknik berbasis *machine*

*learning* yang masih jarang digunakan untuk melakukan proses prediksi, baik digunakan dalam kasus klasifikasi maupun regresi, dan sangat populer belakangan ini. *Support Vector Machine* merupakan satu dari sekian banyak metode yang dapat digunakan untuk menyelesaikan berbagai macam permasalahan diantaranya ialah masalah prediksi. Dalam menyelesaikan sebuah permasalahan, *Support Vector Machine* mampu menangani permasalahan *non-linier* dengan adanya fungsi *kernel* yang membuat metode ini dapat digunakan untuk peramalan *time series* (Kusumodestoni & Sarwido, 2017).

Penelitian yang sama pernah dilakukan oleh Harold Situmorang pada tahun 2015 dalam Tesisnya yang berjudul “Klasifikasi Wilayah Epidemis Demam Berdarah Menggunakan Metode *Support Vector Machine* Di Kota Medan” dalam penelitiannya *variabel* yang digunakan untuk proses analisis klasifikasi adalah tinggih wilayah, curah hujan, kelembapan udara, kepadatan penduduk, jumlah kasus Demam Berdarah Dengue (DBD), dan dalam Proses klasifikasi wilayah Demam berdarah memperoleh akurasi pengujian sebanyak 75 %.

Tujuan dari penelitian ini adalah untuk membangun sebuah model prediksi penyakit DBD di Kota Bandar Lampung dengan menggunakan metode *Support Vector Machine*. Dalam pembangunan model prediksi ini menggunakan 3 *kernel* yaitu linear, gaussian dan polynomial.

## **B. Rumusan Masalah**

Rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Memprediksi jumlah penderita Demam Berdarah di Kota Bandar Lampung menggunakan data *numerik*.
2. Mengukur kinerja dari hasil prediksi jumlah penderita demam berdarah di Kota Bandar Lampung.
3. Membandingkan hasil kinerja prediksi dengan hasil kinerja pada penelitian sebelumnya.

## **C. Tujuan**

Tujuan dalam penelitian ini adalah untuk mengukur kinerja dalam metode *Support Vector Machine* menggunakan *kernel* linier, gaussian dan polynomial serta membandingkan dengan penelitian sebelumnya yang dilakukan oleh Harold Situmorang pada tesisnya yang berjudul Klasifikasi Wilayah Epidemik Demam Berdarah Menggunakan Metode *Support Vector Machine* Di Kota Medan.

## **D. Batasan Masalah**

Adapun batasan masalah dalam penelitian ini adalah sebagai berikut :

1. Penelitian ini hanya berfokus pada pembuatan model SVM untuk membangun model prediksi DBD di Kota Bandar Lampung.
2. Jumlah data yang digunakan sebanyak 30 puskesmas se kota Bandar Lampung dari tahun 2010 sampai 2018.

3. Dalam penelitian ini menggunakan 3 *kernel* yaitu linear, gaussian dan polynomial.

#### **E. Manfaat**

Manfaat dari penelitian ini diharapkan dapat memberikan referensi bagi penelitian selanjutnya yang berkaitan dengan pembangunan model prediksi dengan metode *Support Vector Machine*. Selain itu juga, dalam penelitian ini diharapkan dapat memberikan pengetahuan kepada masyarakat mengenai penyakit demam berdarah krususnya di daerah kota Bandar Lampung dan memberikan bayangan kepada Dinas Kesehatan Kota Bandar Lampung untuk melakukan tindakan pencegahan sejak dini terhadap penyakit Demam Berdarah.

## II. TINJAUAN PUSTAKA

### A. *Machine learning*

*Machine Learning* adalah metode yang digunakan untuk membuat program yang bisa belajar dari data. Berbeda dengan program komputer biasa yang statis, program machine learning adalah program yang dirancang untuk mampu belajar sendiri (W. Yuciana, 2014)

Cara belajar program *machine learning* mengikuti cara belajar manusia, yakni belajar dari contoh-contoh. *Machine learning* akan mempelajari pola dari contoh-contoh yang dianalisa, untuk menentukan jawaban dari pertanyaan-pertanyaan berikutnya. Memang tidak semua masalah bisa dipecahkan dengan program *machine learning*. Namun, seringkali algoritma yang sifatnya kompleks, ternyata bisa dipecahkan dengan sangat simpel oleh *machine learning*. Beberapa contoh program *machine learning* yang telah digunakan dalam kehidupan sehari-hari: Pendeteksi Spam, Pendeteksi Wajah, Rekomendasi Produk, Asisten Virtual, Diagnosa Medis, Pendeteksi Penipuan Kartu Kredit, Pengenal Digit, Perdagangan Saham dan Segmentasi Pelanggan.



## B. Alur Kerja *Machine Learning*

Proses yang dilakukan dalam alur kerja *machine learning* ialah mengumpulkan dataset yang ingin di uji, kemudian memilih metode yang ingin digunakan misalnya *regresi linear*, *regresi logistik*, *neural network*, *svm* dan lain-lain, Memberikan latihan terhadap metode yang dipilih, Melakukan evaluasi terhadap metode yang dipilih dan yang terakhir melakukan Prediksi (Ahmad, 2017)

Akurasi awal dari program *machine learning* biasanya sangat buruk. Namun, seiring berjalannya waktu, semakin sering kita melatih program, semakin banyak contoh-contoh yang dipelajari oleh program, maka program ini akan semakin baik dan akurat.

## C. *Machine Learning* Berdasarkan tipe data

Berdasarkan tipe atau jenisnya *machine learning* di bagi menjadi dua jenis yaitu *Supervised Learning* dan *Unsupervised Learning* dari *Supervised Learning* dibagi lagi menjadi dua yaitu adalah *regression* dan *classification* (Chairani, Widyawan, & Kusumawardani, 2015). Berikut merupakan penjelasan dari tipe dan jenis *machine learning* :

### 1. *Supervised Learning*:

Menurut Chairani pada tahun 2015 *Supervised learning* merupakan suatu teknik pembelajaran dengan membuat suatu fungsi tertentu dari data latihan. Data latihan terdiri dari pasangan nilai input, dan *output*.

Tugas utama dari *supervised learning* adalah memprediksi nilai fungsi untuk semua nilai input yang telah melalui proses pelatihan.

**a. *Regression***

Data yang ada diberikan *real value*, *numerical* atau *floating point*, agar dapat mencoba mendeteksi harga saham di kemudian hari.

Contoh dari penggunaan metode *regression* ini biasanya di gunakan pada data yang bersifat *time series*.

**b. *Classification***

Data yang ada diberikan label atau kategori, agar dapat diambil keputusan berdasarkan label atau kategori tersebut.

**2. *Unsupervised Learning***

Data tidak diberikan label, tapi secara *otomatis* dibagi berdasarkan kemiripan dan struktur lain dari data tersebut. Misalnya, ketika kita mengorganisasikan foto. Kita harus melakukan tagging secara manual

**D. *Cross validation***

**1. *k-fold crossvalidation***

*k-fold crossvalidation* merupakan suatu metode yang biasa digunakan untuk melakukan evaluasi kinerja *classifier*, metode ini dapat digunakan apabila memiliki jumlah data yang sedikit. *k-fold crossvalidation* juga biasanya digunakan untuk mengetahui rata-rata keberhasilan dari suatu sistem dengan cara melakukan sebuah pengulangan dengan cara mengacak atribut masukan sehingga sistem

tersebut teruji untuk beberapa atribut input yang acak yang baru (Banjarsari, Budiman, & Farmadi, 2016).

## **2. Stratified K-Fold Cross Validation**

Dalam beberapa kasus, mungkin ada ketidakseimbangan yang besar dalam variabel respon. Misalnya, dalam dataset tentang harga rumah, mungkin ada sejumlah besar rumah memiliki harga tinggi. Atau dalam hal klasifikasi, mungkin ada beberapa kali sampel negatif lebih banyak daripada sampel positif. Untuk masalah seperti itu, sedikit variasi dalam teknik validasi silang, sehingga setiap lipatan berisi kira-kira persentase sampel yang sama dari setiap kelas target sebagai set lengkap, atau dalam kasus masalah prediksi, nilai respons rata-rata adalah sekitar sama di semua lipatan. Variasi ini juga dikenal sebagai Stratified K Fold. Teknik validasi yang dijelaskan di atas juga disebut sebagai metode validasi silang tidak lengkap. Ini tidak menghitung semua cara untuk memecah sampel asli,. Anda hanya perlu memutuskan berapa banyak subset yang perlu dibuat (Banjarsari, Budiman, & Farmadi, 2016).

## **3. Leave-P-Out Cross Validation**

Pendekatan ini meninggalkan poin data  $p$  dari data pelatihan, yaitu jika ada titik data  $n$  dalam sampel asli maka, sampel  $n-p$  digunakan untuk melatih model dan titik  $p$  digunakan sebagai set validasi. Ini diulangi untuk semua kombinasi di mana sampel asli dapat dipisahkan dengan

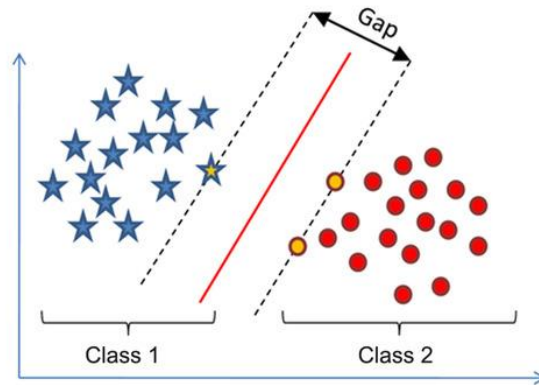
cara ini, dan kemudian kesalahan dirata-rata untuk semua percobaan, untuk memberikan efektivitas keseluruhan. Metode ini lengkap, dalam arti bahwa ia perlu melatih dan memvalidasi model untuk semua kombinasi yang mungkin, dan untuk  $p$  yang cukup besar, dapat menjadi tidak layak secara komputasi. Validasi Silang adalah teknik yang sangat berguna untuk menilai efektivitas model Anda, terutama dalam kasus di mana Anda perlu mengurangi overfitting. Ini juga berguna dalam menentukan parameter hiper model Anda, dalam arti bahwa parameter mana yang akan menghasilkan kesalahan pengujian terendah. Ini semua dasar yang Anda butuhkan untuk memulai dengan validasi silang (Banjarsari, Budiman, & Farmadi, 2016).

#### **E. SVM ( *Support Vector Machine* )**

SVM memiliki kemampuan untuk mengatasi masalah yang bersifat linear dan nonlinear. Pendekatan *kernel* yang dikembangkan pada SVM juga dapat digunakan untuk mengatasi jumlah kelas yang beragam dan banyak, SVM untuk prediksi disebut dengan SVM *Regression* terdiri dari fungsi linear dan nonlinear. Metode SVM *Regression* disebut sebagai suatu teknik *nonparametric* karena bergantung pada fungsi *kernel* (Nurmasani, 2017).

Konsep SVM dapat dijabarkan sebagai suatu cara atau metode untuk mencari *hyperplane* terbaik yang berfungsi sebagai pemisah antar dua buah *class* pada input *space* seperti yang diperlihatkan pada Gambar 1, konsep tersebut memiliki prinsip *Structural Risk Minimization* atau yang biasa

disebut dengan SRM, dengan tujuan menemukan *hyperplane* terbaik yang dapat memisahkan data ke dalam dua buah kelas (Honakan, Adiwijaya, 2018).



Gambar 1. Pembagian data menjadi 2 class pada kernel linear (Cortes dan Vapnik, 1995).

Konsep SVM dapat dijelaskan secara sederhana sebagai usaha mencari *hyperplane* terbaik yang berfungsi sebagai pemisah dua buah kelas pada input space. Pattern yang merupakan anggota dari dua buah kelas : +1 dan -1 dan berbagi alternative garis pemisah (*discrimination boundaries*). *Margin* adalah jarak antara *hyperplane* tersebut dengan *pattern* terdekat dari masing-masing kelas. *Pattern* yang paling dekat ini disebut sebagai *support vector*. Usaha untuk mencari lokasi *hyperplane* ini merupakan inti dari proses pembelajaran pada SVM. Data yang tersedia dinotasikan sebagai  $\vec{x}_i \in \mathbb{R}^d$  sedangkan label masing-masing dinotasikan  $y_i \in \{-1, +1\}$  untuk  $i = 1, 2, \dots, l$ , yang mana  $l$  adalah banyaknya data. Diasumsikan kedua kelas -1 dan +1 dapat terpisah secara sempurna oleh *hyperplane* berdimensi  $d$ , yang didefinisikan :

$$\vec{w} \cdot \vec{x} + b = 0 \dots\dots\dots(1)$$



Pattern  $\vec{x}_i$  yang termasuk kelas -1 (sampel negatif) dapat dirumuskan sebagai *pattern* yang memenuhi pertidaksamaan :

$$\vec{w} \cdot \vec{x}_i + b \leq -1 \dots\dots\dots(2)$$

Sedangkan pattern  $\vec{x}_i$  yang termasuk kelas +1 (sampel positif) memenuhi pertidaksamaan :

$$\vec{w} \cdot \vec{x}_i + b \geq +1 \dots\dots\dots(3)$$

*Margin* terbesar dapat ditemukan dengan memaksimalkan nilai jarak antara *hyperplane* dan titik terdekatnya, yaitu  $1/\|\vec{w}\|$ . Hal ini dapat dirumuskan sebagai *Quadratic Programming* (QP) Problem, yaitu mencari titik minimal persamaan (4) dengan memperhatikan constraint persamaan (8).

$$\min \tau(w) = \frac{1}{2} \|\vec{w}\|^2 \dots\dots\dots(4)$$

$$y_i (x_i \cdot w + b) - 1 \geq 0, \forall i \dots\dots\dots(5)$$

Problem ini dapat dipecahkan dengan berbagai teknik komputasi, diantaranya dengan *Lagrange Multiplier*.

$$L(w, b, \alpha) = \frac{1}{2} \|\vec{w}\|^2 - \sum_{i=1}^l \alpha_i (y_i ((x_i \cdot w) + b) - 1) \dots\dots\dots(6)$$

$\alpha_i$  adalah *Lagrange Multiplier*, yang bernilai nol atau positif ( $\alpha_i \geq 0$ ). Nilai optimal dari persamaan (6) dapat dihitung dengan meminimalkan  $L$  terhadap  $\vec{w}$  dan  $b$ , dan memaksimalkan  $L$  terhadap  $\alpha_i$ . Dengan memperhatikan sifat bahwa pada titik optimal gradient  $L = 0$ , persamaan (6)

dapat dimodifikasi sebagai maksimalisasi problem yang hanya mengandung  $\alpha_i$ , sebagaimana persamaan (7) berikut.

Maksimasi :

$$\sum_{i=1}^l \alpha_i - \sum_{j=1}^l \alpha_j y_i y_j x_i x_j \dots \dots \dots (7)$$

Dengan Constraint :

$$\alpha_i \geq 0 \ (i = 1, 2, 3, \dots, l) \ \sum_{i=1}^l \alpha_i \dots \dots \dots (8)$$

Dari perhitungan ini diperoleh  $\alpha_i$  yang kebanyakan bernilai positif. Data yang berkorelasi dengan  $\alpha_i$  yang positif inilah yang disebut *support vector*.

Penjelasan diatas berdasarkan asumsi bahwa kedua belah kelas dapat terpisah secara sempurna oleh *hyperplane*. Akan tetapi, pada umumnya dua belah kelas pada *input space* tidak dapat terpisah secara sempurna (*non linear separable*). Hal ini menyebabkan *constraint* pada persamaan (8) tidak dapat terpenuhi, sehingga optimisasi tidak dapat dilakukan. Untuk mengatasi masalah ini, SVM dirumuskan dengan memperkenalkan teknik *softmargin*.

Dalam *softmargin*, persamaan (5) dimodifikasi dengan memasukkan *slack variable*  $\zeta_i$  ( $\zeta > 0$ ) sebagai berikut :

$$y_i (x_i \cdot w + b) \geq 1 - \zeta_i, \forall i \dots \dots \dots (9)$$

dengan demikian persamaan (4) diubah menjadi

$$\min \tau(w) = \frac{1}{2} \|\bar{w}\|^2 + C \sum_{i=1}^l \zeta_i \dots \dots \dots (10)$$

Parameter C dipilih untuk mengontrol *tradeoff* antara margin dan error klasifikasi  $\zeta$ . Nilai C yang besar berarti akan memberikan *penalty* yang lebih besar terhadap error klasifikasi tersebut.

Untuk menyelesaikan problem *non-linear*, SVM dimodifikasi dengan memasukkan fungsi kernel. Dalam *non-linear* SVM, karena umumnya transformasi ini tidak diketahui, dan sangat sulit untuk difahami secara mudah, maka perhitungan *dot product* dapat digantikan dengan fungsi kernel  $K(\vec{x}_i \cdot \vec{x}_j)$  yang mendefinisikan secara implisit. Hal ini disebut sebagai Kernel Trick, yang dirumuskan :

$$K(\vec{x}_i \cdot \vec{x}_j) = (\vec{x}_i) \cdot (\vec{x}_j) \dots\dots\dots(11)$$

$$f(\vec{x}_i) = \vec{w} \cdot (\vec{x}_i) + b \dots\dots\dots(12)$$

$$= \sum_{i=1}^l, x_i \in SV \alpha_i y_i (\vec{x}) \cdot (\vec{x}_i) + b \dots\dots\dots(13)$$

$$= \sum_{i=1}^l, x_i \in SV \alpha_i y_i K(x, x_i) + b \dots\dots\dots(14)$$

Syarat sebuah fungsi untuk menjadi fungsi kernel adalah memenuhi teorema Mercer yang menyatakan bahwa matriks kernel yang dihasilkan harus bersifat *positive semi-definite*. Fungsi kernel yang umum digunakan adalah sebagai berikut:

a. Kernel Linear

$$K(x_i, x_j) = x_i \cdot x_j + C \dots\dots\dots(15)$$

b. Kernel Gaussian

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2), \gamma > 0 \dots\dots\dots(16)$$

## c. Kernel Polynomial

$$K(x_i, x_j) = (\gamma x_i \cdot x_j + r)^d, \gamma > 0 \dots \dots \dots (17)$$

## d. Sigmoid

$$K(\vec{x}, \vec{y}) = \tanh(K < \vec{x} \cdot \vec{y} > + \Theta) \dots \dots \dots (18)$$

Pada penelitian ini kernel yang digunakan adalah kernel linier, gaussian dan polynomial. Dengan memasukkan fungsi kernel pada persamaan optimasi SVM, maka persamaan menjadi

$$\max \{ \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j K(\vec{x}_i, \vec{x}_j) \}$$

$$\begin{aligned} s.t \quad & \sum_{i=1}^l \alpha_i y_i = 0 \\ & 0 \leq \alpha_i \leq C \quad i = 1, \dots, n \\ & \beta_d \geq 0 \quad i = 1, \dots, m \end{aligned}$$

1. **Kernel SVM Linear**

*Kernel* Linear merupakan fungsi *kernel* paling sederhana (Bhavsar & Panchal, 2012). *Kernel* ini sering digunakan pada data set yang sebaran datanya dapat diklasifikasikan dengan cara linear. Fungsi *kernel* SVM Linear dapat dilihat pada persamaan 19.

$$K(x_i, x_j) = x_i \cdot x_j + C \dots \dots \dots (19)$$

Dengan  $x_i$  dan  $x_j$  merupakan vektor dari data set dan  $C$  merupakan constant. Dalam Penelitian ini saya menggunakan *library* dari kernel Linear, yaitu metodeSVM = "svmLinear", Jika diuraikan *library* kernel linear akan terlihat seperti pada Kode Program 1.

```
# parameters needed for linear kernel
if (kernel.function == "linear" & !is.null(rho)) {
  rho <- NULL
```

```

    warning("rho = ", call.names[call.order[4]], " is not used with linear
kernel")
  }
  if (kernel.function == "linear" & !is.null(gamma)) {
    gamma <- NULL
    warning("gamma = ", call.names[call.order[5]], " is not used with linear
kernel")
  }
  if (kernel.function == "linear" & !is.null(d)) {
    d <- NULL
    warning("d = ", call.names[call.order[6]], " is not used with linear
kernel")
  }
}

```

Kode Perogram 1. Code Implementasi Kernel Linear Ke R Studio.

## 2. Kernel SVM Gaussian

*Kernel SVM Gaussian* merupakan *kernel* pada SVM yang digunakan untuk menyelesaikan masalah yang tidak bisa diselesaikan dengan cara Linear (Bhavsar & Panchal, 2012). Fungsi *kernel SVM Gaussian* dapat dilihat pada persamaan 20.

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2), \gamma > 0 \dots\dots\dots (20)$$

Dengan  $x_i$  dan  $x_j$  merupakan vektor dari data set,  $\gamma$  adalah parameter untuk mengkontrol kecepatan proses learning dan  $\exp$  merupakan basis dari logaritma alami. Dalam Penelitian ini saya menggunakan *library* dari kernel Gaussian, yaitu `metodeSVM = "svmRadial"`, Jika diuraikan *library* kernel gaussian akan terlihat seperti pada Kode Program 2.

```

# parameters needed for gaussian kernel
if (kernel.function == "gaussian" & !is.null(rho)) {
  if (rho <= 0) {
    stop("rho should be positive")
  }
}

```

```

}
if (kernel.function == "gaussian" & !is.null(gamma)) {
  gamma <- NULL
  warning("gamma = ", call.names[call.order[5]], " is not used with
gaussian kernel")
}
if (kernel.function == "gaussian" & !is.null(d)) {
  d <- NULL

  warning("d = ", call.names[call.order[6]], " is not used with gaussian
kernel")

```

Kode Perogram 2. Code Implementasi Kernel Gaussian Ke R Studio.

### 3. *Kernel SVM Polynomial*

Persamaan Polynomial merupakan sebuah persamaan yang terdiri dari variabel dan koefisien yang memiliki suku banyak (Bhavsar & Panchal, 2012). *Kernel SVM Polynomial* sering digunakan dalam masalah dimana semua *training data*-nya dinormalisasi. Fungsi *kernel SVM Polynomial* dapat dilihat pada persamaan 21.

$$K(x_i, x_j) = (\gamma x_i \cdot x_j + r)^d, \gamma > 0 \dots \dots \dots (21)$$

Dengan  $x_i$  dan  $x_j$  merupakan vektor dari data set,  $\gamma$  adalah parameter untuk mengontrol kecepatan proses learning dan  $d$  merupakan pangkat polynomial yang digunakan. Dalam Penelitian ini saya menggunakan *library* dari kernel Polynomial, yaitu metodeSVM = "svmPoly", Jika diuraikan *library* kernel Polynomial akan terlihat seperti pada Kode Program 21.

```

# parameters needed for polynomial kernel
if (kernel.function == "polynomial" & !is.null(rho)) {
  if (rho <= 0) {
    stop("rho should be positive")
  }
}

```

```

if (kernel.function == "polynomial" & !is.null(d)) {
  if (kernel.function == "polynomial" & (d <= 0 | !check.integer(d))) {
    stop("d should be an integer > 0")
  }
}
if (kernel.function == "polynomial" & is.null(d)) {
  stop("d is missing")
}

```

Kode Perogram 3. Code Implementasi Kernel Polynomial Ke R Studio.

## F. Uji Koefisien Determinasi

Untuk mengetahui ketepatan atau kecocokan garis regresi yang terbentuk dalam mewakili kelompok data hasil observasi, perlu dilihat sampai seberapa jauh model yang terbentuk mampu menerangkan kondisi yang sebenarnya yang dikenal dengan nama koefisien determinasi ( $R^2$ ). Nilai koefisien determinasi merupakan suatu ukuran yang menunjukkan besar sumbangan dari variabel penjelasan terhadap respon (Siagaan, 2006).

### 1. *Sum Of Squares Error (SSE)*

Sum Of Squares Error (SSE) adalah mengukur kesalahan penggunaan estimasi persamaan regresi untuk menghitung nilai variabel terikat dari sampel. Fungsi dari SSE dapat dilihat pada persamaan 4.

$$SSE = \sum (Y_i - \hat{Y}_i)^2 \dots\dots\dots(4)$$

Dengan  $Y_i$  merupakan data aktual dan  $\hat{Y}_i$  merupakan data hasil prediksi.

### 2. *Total Sum Of Squares (TSS)*

Total sum of Squares (TSS) adalah menggunakan nilai rata-rata (mean) hasil penjualan kuartalan dari sampel. Fungsi dari TSS dapat dilihat pada persamaan 5.

$$SST = \sum (Y_i - \hat{Y})^2 \dots\dots\dots(5)$$

Dengan  $Y_i$  merupakan data aktual dan  $\hat{Y}$  merupakan data *testing*.

### 3. R-squared ( $R^2$ )

R-squared ( $R^2$ ) adalah suatu indikator yang digunakan untuk menggambarkan berapa banyak variasi yang dijelaskan dalam model. Berdasarkan nilai  $R^2$  dapat diketahui tingkat signifikansi atau kesesuaian hubungan antara *variabel* bebas dan *variabel* tak bebas dalam regresi.  $R^2$  dengan nilai mendekati 1 berarti model regresi yang telah kita buat cukup baik dalam memprediksi data dan jika nilai  $R^2$  mendekati 0 maka model yang kita buat kurang baik dalam memprediksi data. Fungsi dari R-squared ( $R^2$ ) dapat dilihat pada persamaan 6.

$$R^2 = 1 - SSE/SST \dots\dots\dots(6)$$



### **III. METODOLOGI PENELITIAN**

#### **A. Waktu dan Tempat Penelitian**

##### **1. Tempat Penelitian**

Penelitian ini dilaksanakan di Jurusan Ilmu Komputer Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung yang beralamatkan di jalan Soemantri Brojonegoro No. 1 Gedung Meneng, Bandar Lampung.

##### **2. Waktu penelitian**

Penelitian ini dilaksanakan pada semester ganjil tahun ajaran 2018/2019, Pada minggu 1 sampai 4 bulan Desember peneliti mulai mengumpulkan data, pada minggu 1 sampai 4 di bulan Januari dimulai proses pemodelan data untuk diprediksi, pada minggu ke 1 Januari sampai minggu ke 4 bulan Februari dilakukan coding, pada minggu ke 1 Maret sampai minggu ke 4 di bulan Mei dilakukan Pencarian akurasi terbaik untuk hasil prediksinya, pada bulan Juni dilakukan penyusunan laporan dan penarikan kesimpulan.

## B. Data dan Alat

### 1. Data

Data yang di gunakan pada penelitian ini adalah data penyakit demam berdarah di kota Bandar Lampung dari tahun 2010 sampai dengan tahun 2018, data tersebut diambil dari Dinas kesehatan Kota Bandar Lampung. Pada Tabel 2 menampilkan data penderita DBD pada tahun 2018 data tersebut merupakan gambaran untuk data yang digunakan pada proses prediksi nantinya.

#### a. *Feature* Dalam Data

- *Latitude* dan *Longitude* adalah suatu sistem koordinat geografis yang digunakan untuk menentukan lokasi suatu tempat di permukaan bumi, data ini diperoleh dari *google maps*.
- Curah hujan adalah jumlah air yang jatuh di permukaan tanah datar selama periode tertentu yang diukur dengan satuan tinggi (mm). Dalam penelitian ini data curah hujan dibagi menjadi 3, yaitu : curah hujan maksimal, minimal dan rata-rata. Data ini diperoleh dari *website* Badan Pusat Statistika (BPS) di alamat <https://lampung.bps.go.id/dynamictable/2016/11/23/71/jumlah-curah-hujan-provinsi-lampung-2010-2018.html>
- *Temperatur* adalah ukuran panas-dinginnya dari suatu benda. Panas-dinginnya suatu benda berkaitan dengan energi termis yang terkandung dalam benda tersebut. Makin besar energi termisnya, makin besar temperaturnya, dengan satuan (C°). Dalam penelitian ini data temperatur dibagi menjadi 3, yaitu :

Temperatur maksimal, minimal dan rata-rata. Data ini diperoleh dari *website* Badan Pusat Statistika (BPS) dengan alamat <https://www.bps.go.id/statictable/2014/04/28/1347/suhuminimum-rata-rata-dan-maksimum-di-stasiun-pengamatan-bmkg-oc-2000-2018.html>.

- Kelembapan udara adalah jumlah kandungan uap air yang ada dalam udara. Kandungan uap air di udara berubah-ubah bergantung suhu udara, dengan satuan (%). Dalam penelitian ini data kelembapan udara dibagi menjadi 3, yaitu : kelembapan udara maksimal, minimal dan rata-rata. Data ini diperoleh dari *website* Badan Pusat Statistika (BPS) dengan alamat <https://lampung.bps.go.id/dynamictable/2017/03/27/108/ratarata-kelembaban-udara-di-provinsi-lampung-2010-2018.html>.
- Kecepatan angin, atau *velositas* gelombang angin, adalah sebuah kuantitas atmosferik fundamental. Kecepatan angin disebabkan oleh pergerakan angin dari tekanan tinggi ke tekanan rendah, dengan satuan (Knot). Dalam penelitian ini data kecepatan angin dibagi menjadi 3, yaitu : Kecepatan angin maksimal, minimal dan rata-rata. Data ini diperoleh dari *website* Badan Pusat Statistika (BPS) dengan alamat *website* <https://lampung.bps.go.id/dynamictable/2017/03/27/107/ratarata-kecepatan-angin-di-provinsi-lampung-2010-2018.html>.
- Jumlah penduduk adalah jumlah manusia yang bertempat tinggal atau berdomisili pada suatu wilayah atau daerah dan

memiliki mata pencaharian tetap. Data ini diperoleh dari *website* Badan Pusat Statistika (BPS) dengan alamat *website* <https://bandarlampungkota.bps.go.id/statictable/2015/12/08/19/jumlah-penduduk-kota-bandar-lampung-2010-2020.html>.

- Luas area adalah besaran yang menyatakan ukuran dua dimensi suatu bagian permukaan yang dibatasi dengan jelas, dengan satuan ( $\text{Km}^2$ ). Data ini dari *website* BPS dengan alamat <https://bandarlampungkota.bps.go.id/statictable/2015/12/08/9/luas-wilayah-kota-bandar-lampung-menurut-kecamatan-tahun-2014-km2-.html>.
  - Matriks jarak adalah jarak lokasi antar puskesmas yang satu dengan puskesmas yang lainnya, dengan satuan ( $\text{Km}^2$ ) Data ini diperoleh dari perhitungan *Latitude* dan *Longitude*. Dalam fungsi penghitungan jarak berikut, dengan menggunakan Micrisoft Excel dengan rumus pada persamaan 7.
- $$=6370*ACOS(SIN(J19)*SIN(M19)+COS(J19)*COS(M19)*COS(M21-J21)) \dots\dots\dots(7)$$

Tabel 1. Contoh Data DBD Puskesmas Bakung (Laporan DBD Tahunan Dinas Kesehatan Kota Bandar Lampung).

| No | puskes | Thn  | Bulan | Penderita | L | P | Jmlh | latitude   | longitude   | temMin | temMax | Jmlh<br>pnddk | Area (km <sup>2</sup> ) |
|----|--------|------|-------|-----------|---|---|------|------------|-------------|--------|--------|---------------|-------------------------|
| 1  | Bakung | 2010 | 3     | 1         | 0 | 1 | 8    | -5.454.793 | 105.248.137 | 23,2   | 33,6   | 29.799        | 11,02                   |
| 2  |        |      | 6     | 1         | 0 | 1 |      | -5.454.793 | 105.248.137 | 23,5   | 32,4   |               | 11,02                   |
| 3  |        |      | 9     | 4         | 1 | 3 |      | -5.454.793 | 105.248.137 | 24,1   | 33,4   |               | 11,02                   |
| 4  |        |      | 12    | 2         | 1 | 1 |      | -5.454.793 | 105.248.137 | 25,6   | 31,2   |               | 11,02                   |
| 5  |        | 2011 | 3     | 1         | 1 | 0 | 7    | -5.454.793 | 105.248.137 | 21,0   | 34,3   | 30.000        | 11,02                   |
| 6  |        |      | 6     | 1         | 1 | 0 |      | -5.454.793 | 105.248.137 | 21,5   | 31,2   |               | 11,02                   |
| 7  |        |      | 9     | 4         | 3 | 1 |      | -5.454.793 | 105.248.137 | 22,6   | 34,3   |               | 11,02                   |
| 8  |        |      | 12    | 1         | 1 | 0 |      | -5.454.793 | 105.248.137 | 25,8   | 34,1   |               | 11,02                   |
| 9  |        | 2012 | 3     | 1         | 1 | 0 | 9    | -5.454.793 | 105.248.137 | 21,2   | 34,1   | 30.775        | 11,02                   |
| 10 |        |      | 6     | 3         | 1 | 2 |      | -5.454.793 | 105.248.137 | 21,8   | 32,4   |               | 11,02                   |
| 11 |        |      | 9     | 4         | 2 | 2 |      | -5.454.793 | 105.248.137 | 22,6   | 31,2   |               | 11,02                   |
| 12 |        |      | 12    | 1         | 1 | 0 |      | -5.454.793 | 105.248.137 | 23,4   | 30,2   |               | 11,02                   |
| 13 |        | 2013 | 3     | 1         | 1 | 0 | 11   | -5.454.793 | 105.248.137 | 22,9   | 33,3   | 31.500        | 11,02                   |
| 14 |        |      | 6     | 4         | 3 | 1 |      | -5.454.793 | 105.248.137 | 23,1   | 33,1   |               | 11,02                   |
| 15 |        |      | 9     | 5         | 2 | 3 |      | -5.454.793 | 105.248.137 | 25,7   | 31,2   |               | 11,02                   |
| 16 |        |      | 12    | 1         | 1 | 0 |      | -5.454.793 | 105.248.137 | 28,1   | 30,9   |               | 11,02                   |
| 17 |        | 2014 | 3     | 1         | 0 | 1 | 15   | -5.454.793 | 105.248.137 | 23,5   | 34,5   | 32.023        | 11,02                   |
| 18 |        |      | 6     | 4         | 2 | 2 |      | -5.454.793 | 105.248.137 | 23,6   | 34,1   |               | 11,02                   |
| 19 |        |      | 9     | 8         | 4 | 4 |      | -5.454.793 | 105.248.137 | 26,1   | 33,2   |               | 11,02                   |
| 20 |        |      | 12    | 2         | 2 | 0 |      | -5.454.793 | 105.248.137 | 27,3   | 30,8   |               | 11,02                   |
| 21 |        | 2015 | 3     | 3         | 3 | 0 | 18   | -5.454.793 | 105.248.137 | 20,2   | 37,2   | 32.808        | 11,02                   |
| 22 |        |      | 6     | 5         | 3 | 2 |      | -5.454.793 | 105.248.137 | 21,3   | 36,3   |               | 11,02                   |

|    |  |      |    |    |   |   |    |            |             |       |      |        |       |
|----|--|------|----|----|---|---|----|------------|-------------|-------|------|--------|-------|
| 23 |  |      | 9  | 9  | 5 | 4 |    | -5.454.793 | 105.248.137 | 28,7  | 37,1 |        | 11,02 |
| 24 |  |      | 12 | 1  | 1 | 0 |    | -5.454.793 | 105.248.137 | 27,2  | 35,2 |        | 11,02 |
| 25 |  | 2016 | 3  | 3  | 2 | 1 | 20 | -5.454.793 | 105.248.137 | 20,8  | 34,6 | 33.043 | 11,02 |
| 26 |  |      | 6  | 4  | 2 | 2 |    | -5.454.793 | 105.248.137 | 21,6  | 33,5 |        | 11,02 |
| 27 |  |      | 9  | 11 | 8 | 3 |    | -5.454.793 | 105.248.137 | 24,6  | 32,9 |        | 11,02 |
| 28 |  |      | 12 | 2  | 1 | 1 |    | -5.454.793 | 105.248.137 | 28,4  | 30,1 |        | 11,02 |
| 29 |  | 2017 | 3  | 2  | 1 | 1 | 25 | -5.454.793 | 105.248.137 | 21,30 | 34,0 | 33.964 | 11,02 |
| 30 |  |      | 6  | 5  | 4 | 1 |    | -5.454.793 | 105.248.137 | 23,6  | 32,9 |        | 11,02 |
| 31 |  |      | 9  | 12 | 6 | 6 |    | -5.454.793 | 105.248.137 | 21,2  | 31,4 |        | 11,02 |
| 32 |  |      | 12 | 6  | 2 | 4 |    | -5.454.793 | 105.248.137 | 29,6  | 30,8 |        | 11,02 |
| 33 |  | 2018 | 3  | 3  | 3 | 0 | 24 | -5.454.793 | 105.248.137 | 21,30 | 33,0 | 34.143 | 11,02 |
| 34 |  |      | 6  | 6  | 0 | 6 |    | -5.454.793 | 105.248.137 | 23,1  | 32,9 |        | 11,02 |
| 35 |  |      | 9  | 12 | 6 | 6 |    | -5.454.793 | 105.248.137 | 25,6  | 31,6 |        | 11,02 |
| 36 |  |      | 12 | 3  | 3 | 0 |    | -5.454.793 | 105.248.137 | 28,1  | 32,8 |        | 11,02 |

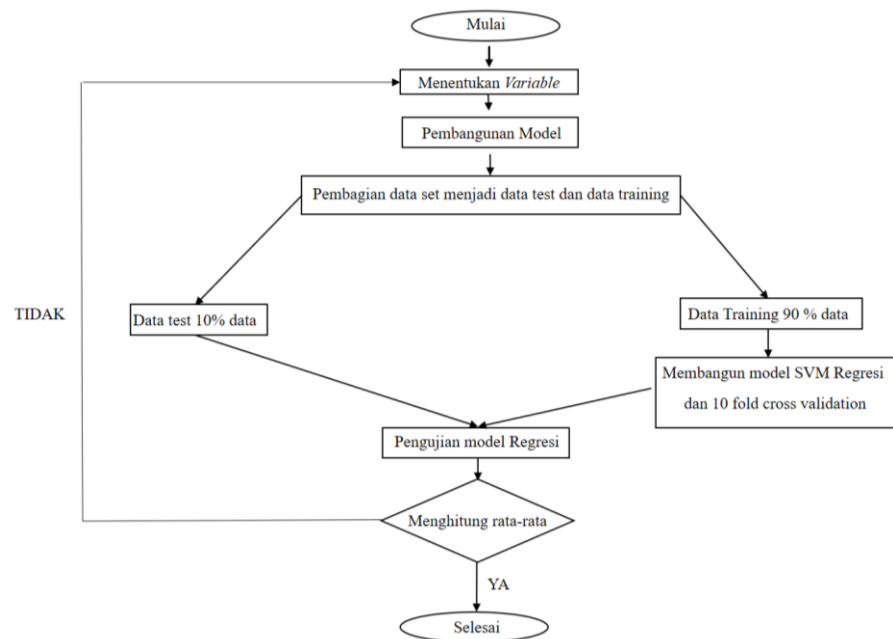
## 2. Alat

Dalam penelitian ini peneliti menggunakan *Software* dan *Hardware* sebagai berikut :

- *Software* : Sistem operasi Windows 10 Profesional, Microsoft Excel 2010, R studio versi 3.5.1, *Package caret* versi 6.0-81, *Package kernlab* versi 0.9-27 dan R versi 3.5.1.
- *Hardware* : Prosesor Intel® Core™ i5-7200U (2.0 GHz), 4 Threads, 2 Cores, 3 Cache, 4 GT/s OPI Bus Speed, 2.5-3.1GHz CPU Clock Speed, VGA NVIDIA® Geforce® 940Mx, DDR4-2133, RAM 8GB, Memory 128 GB SSD + 1TB HDD, Max Memory Bandwidth 34.1 GB/s.

### C. Metode implementasi

Pada Gambar 2 menampilkan bagaimana cara untuk melakukan prediksi dengan menggunakan metode SVM



Gambar 2. Proses prediksi dengan model svm.

Berdasarkan pada diagram alur metodologi yang ditampilkan pada Gambar 2 dibawah ini merupakan penjelasan dari setiap prosesnya yang akan dilakukan.

### 1. Pengumpulan Data

Pada tahap ini dilakukan proses pengumpulan data yang akan digunakan dalam pembuatan sistem prediksi demam berdarah. Data yang digunakan adalah data penderita DBD dari tahun 2013 sampai 2017 . Data tersebut diperoleh dari dinas kesehatan Kota Bandar Lampung, Selain menggunakan data DBD penelitian ini juga menggunakan data cuaca, temperatur, curah hujan, suhu dan kelembapan udara, data tersebut di ambil dari *website* BPS. Selain menggunakan data DBD dan data cuaca penelitian ini juga menggunakan data jarak yang didapatkan dari perhitungan jarak menggunakan *Latitude* dan *Longitude*.



## 2. Pemodelan dengan SVM

Pada tahap ini mulai dilakukan training SVM pada data latih dengan proses pembelajaran untuk model *Support Vector Machine*. Hal yang terlebih dahulu dilakukan yaitu dengan menentukan tipe *kernel* dan nilai parameter. Jika model terbaik pada data telah terpilih maka sudah bisa untuk dilakukan proses prediksi.

## 3. Pengujian Model

Tahap ini dilakukan ketika sudah mendapatkan model SVM terbaik. Model SVM tersebut akan digunakan untuk melakukan pengujian dengan data testing yang telah kita miliki dari proses pembagian dataset, *k-fold cross validation* merupakan suatu teknik yang dapat digunakan apabila memiliki jumlah data yang terbatas (jumlah instance tidak banyak) *k-fold cross validation* merupakan salah satu metode yang digunakan untuk mengetahui rata-rata keberhasilan dari suatu sistem dengan cara melakukan perulangan dengan mengacak atribut masukan sehingga hasil dari prediksi yang kita lakukan teruji untuk beberapa atribut input yang acak, seperti yang di tampilkan pada Gambar 3.



## V. SIMPULAN DAN SARAN

### A. Simpulan

Dari hasil penelitian yang dilakukan, dapat diambil simpulan sebagai berikut:

1. Percobaan tanpa melakukan *feature selection* didapatkan tingkat akurasi sebesar 70,00% untuk *kernel* Linear, 70,79% untuk *kernel* Gaussian dan 70,61% untuk *kernel* Polynomial. Untuk percobaan tanpa matriks jarak. Akurasi sebesar 74,40% untuk *kernel* Linear, 75,44% untuk *kernel* Gaussian dan 75,10% untuk *kernel* Polynomial. Untuk Percobaan dengan menggunakan matriks jarak.
2. Percobaan dengan melakukan *feature selection* didapatkan akurasi sebesar 74,12% untuk *kernel* Linear, 74,62% untuk *kernel* Gaussian dan 74,58% untuk *kernel* Polynomial. Untuk percobaan tanpa matriks jarak dan akurasi sebesar 74,61% untuk *kernel* Linear, 75,51% untuk *kernel* Gaussian dan 75,15% untuk *kernel* Polynomial. Untuk percobaan dengan menggunakan matriks jarak.
3. Menggunakan data matriks jarak dapat meningkatkan akurasi pada setiap *kernel*

## B. Saran

Adapun saran yang di berikan dalam penelitian ini adalah sebagai berikut

1. Disarankan untuk penelitian selanjutnya dapat menggunakan prangkat yang memiliki spesifikasi lebih tinggi guna memperlancar proses pembuatan model dan proses prediksi pada *study* kasus demam berdarah.
2. Untuk penelitian lebih lanjut dapat dilakukan dengan penambahan beberapa *variable* guna menambah tingkat akurasi, agar mendapatkan nilai akurasi lebih baik lagi.
3. Penelitaian selanjutnya dapat menggunakan metode yang lain untuk dibandingkan.

## DAFTAR PUSTAKA

- Ahmad, A. (2017). Mengenal Artificial Intelligence, Machine Learning, Neural Network, dan Deep Learning. *Teknologi Indonesia*, (June), 1–3.
- Banjarsari, M. A., Budiman, I., & Farmadi, A. (2016). Penerapan K-Optimal Pada Algoritma KNN Untuk Prediksi Kelulusan Tepat Waktu Mahasiswa Program Studi Ilmu Komputer Fmipa Unlam Berdasarkan IP Sampai Dengan Semester 4.
- Bhavsar, H., & Panchal, M. H. (2012). A Review on Support Vector Machine for Data Classification, *I*(10), 185–189.
- Badan Pusat Statistika . Rara - rata Kelembapan Udara Di Provinsi Lampung 2010 – 2019. <https://lampung.bps.go.id/dynamictable/2017/03/27/108/ratarata-kelembapan-udara-di-provinsi-lampung-2010-2018.html>. Diakses pada 03 Maret 2019.
- Badan Pusat Statistika. Jumlah Curah Hujan Provinsi Lampung 2018 - 2019. <https://lampung.bps.go.id/dynamictable/2016/11/23/71/jumlah-curah-hujan-provinsi-lampung-2010-2018.html>. Diakses pada 03 Maret 2019.
- Badan Pusat Statistika. Rata-rata Temperatur Rata – rata dan maksimum. <https://www.bps.go.id/statictable/2014/04/28/1347/suhuminimum-rata-rata-dan-maksimum-di-stasiun-pengamatan-bmkg-oc-2000-2018.html>. Diakses pada 03 Maret 2019.
- Badan Pusat Statistika. Rata-rata Kecepatan Angin Di Provinsi Lampung. <https://lampung.bps.go.id/dynamictable/2017/03/27/107/ratarata-kecepatan-angin-di-provinsi-lampung-2010-2018.html>. Diakses pada 03 Maret 2019.
- Badan Pusat Statistika. Jumlah Penduduk Bandar Lampung 2010 - 2020. <https://bandarlampungkota.bps.go.id/statictable/2015/12/08/19/jumlah-penduduk-kota-bandar-lampung-2010-2020.html>. Diakses pada 03 Maret 2019.

- Chairani, C., Widyawan, W., & Kusumawardani, S. S. (2015). Machine Learning Untuk Estimasi Posisi Objek Berbasis RSS Fingerprint Menggunakan IEEE 802.11g Pada Lantai 3 Gedung JTETI UGM (7), 1-1
- Cortes, C. dan V. Vapnik. (1995). Support-Vector Networks. Machine Learning. The Netherlands. Vol. 20:273-297.
- Fanany, D. (2012). Dengue hemorrhagic fever and natural disaster: the case of Padang, West Sumatra. *International Journal of Collaborative Research on Internal Medicine and Public Health (IJCRIMPH)*, 4(5), 673–678.
- Fauzi, Nungsiyati, Noviarti, T., Muslihudin, M., Irviani, R., & Maseleno, A. (2018). Optimal Dengue Endemic Region Prediction using Fuzzy Simple Additive Weighting based Algorithm. *International Journal of Pure and Applied Mathematics*, 118(7 Special Issue), 473–477.
- Honakan, Adiwijaya, S. A. F. (2018). Analisis Dan Implementasi Support Vector Machine Dengan String *Kernel* Dalam Melakukan Klasifikasi Berita Berbahasa Indonesia Analysis and Implementation Support Vector Machine With String *Kernel* for Classification indonesian news, 5(1), 1701–1710.
- Kusumodestoni, R. H., & Sarwido, S. (2017). Komparasi Model Support Vector Machines (Svm) Dan Neural Network Untuk Mengetahui Tingkat Akurasi Prediksi Tertinggi Harga Saham. *Jurnal Informatika Upgris*, 3(1).
- Nurmasani, A. (2017). Analisis Support Vector Machine. *Jurnal Informasi Interaktif*, 2(1).
- Putro, B., Furqon, M. T., & Wijoyo, S. H. (2018). Prediksi Jumlah Kebutuhan Pemakaian Air Menggunakan Metode Exponential Smoothing ( Studi Kasus : PDAM Kota Malang ), 2(11), 4679–4686.
- Reitermanova, Z. 2010. Data Splitting. WDS'10 Proceedings of Contributed Papers. Czech Republic. Part 1:31-36.
- Setyawan, E., Subantoro, R., & Prabowo, R. (2016). Analisis Peramalan (Forecasting) Produksi Karet (Hevea Brasiliensis) Di PT Perkebunan Nusantara IX Kebun Sukamangli Kabupaten Kendal. *Analisis Peramalan (Forecasting)*, 12(2), 11–19.

- Siagaan, D., Sugiarto. (2006). *Metode Statistika Untuk Bisnis dan Ekonomi*. PT Gramedia Pustaka Utama. Jakarta.
- W. Yuciana. (2014). Penerapan Metode Klasifikasi Support Vector Machine (Svm) Pada Data Akreditasi Sekolah Dasar (SD) Di Kabupaten Magelang, 3(8), 811–820.