

**PERBANDINGAN HASIL METODE *NAÏVE BAYES*
*CLASSIFIER DAN SUPPORT VECTOR MACHINE DALAM
KLASIFIKASI CURAH HUJAN***

(Studi Kasus : Curah Hujan di Jawa Timur Tahun 2013 - 2017)

TUGAS AKHIR

**Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh Gelar Sarjana
Jurusan Statistika**



Riza Indriani Rakhmalia

14611168

**PROGRAM STUDI STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS ISLAM INDONESIA
YOGYAKARTA
2018**

HALAMAN PERSETUJUAN PEMBIMBING

TUGAS AKHIR

Judul : Perbandingan Hasil Metode *Naïve Bayes Classifier* dan *Support Vector Machine* Dalam Klasifikasi Curah Hujan (Studi Kasus : Curah Hujan di Jawa Timur Tahun 2013 - 2017)

Nama Mahasiswa : Riza Indriani Rakhmalia

Nomor Mahasiswa : 14611168

**TUGAS AKHIR INI TELAH DIPERIKSA DAN DISETUJUI UNTUK
DIUJIKAN**

Yogyakarta, Maret 2018

Pembimbing


(Ayundyah Kesumawati, S.Si, M.Si)

HALAMAN PENGESAHAN

TUGAS AKHIR

PERBANDINGAN HASIL METODE *NAÏVE BAYES CLASSIFIER DAN SUPPORT VECTOR MACHINE DALAM KLASIFIKASI CURAH HUJAN*

(Studi Kasus : Curah Hujan di Jawa Timur Tahun 2013 - 2017)



HALAMAN PERSEMBAHAN

Skripsi ini ananda persembahkan untuk :

Bapak dan Ibu yang selalu memberikan arahan, dukungan, semangat, nasihat, motivasi, serta doa setiap waktu, sehingga penelitian ini dapat terselesaikan dengan baik.

Kakak-kakak dan adik-adik peneliti yang selalu bersedia memberi saran untuk masalah-masalah yang ada serta memberikan semangat dan motivasi.

KATA PENGANTAR



Assalamualaikum Wr.Wb

Puji syukur penulis panjatkan atas kehadiran Allah SWT karena atas berkat, rahmat, kesehatan dan kekuatan yang diberikan oleh-nya tugas akhir ini dapat berjalan dengan lancar. Tugas akhir ini memberikan begitu banyak pembelajaran yang kemudian dapat dikembangkan pada penelitian selanjutnya.

Keberhasilan pembuatan tugas akhir ini tentunya tidak terlepas dari pihak-pihak yang membantu, memberi semangat dan turut memberi dukungannya selama kegiatan ini berlangsung. Ucapan terima kasih ini disampaikan kepada :

1. Bapak Drs. Allwar, M.Sc, Ph.D selaku dekan Fakultas Matematika dan Ilmu Pengetahuan Alam.
2. Bapak Dr. RB. Fajriya Hakim, M.Si selaku Ketua Jurusan Statistika beserta jajarannya.
3. Ibu Ayundyah Kesumawati M.Si selaku Dosen Pembimbing penelitian yang telah sabar membimbing.
4. Seluruh pihak yang telah turut mendukung dan membantu jalannya penelitian ini.

Semoga segala bentuk kegiatan yang dilakukan selama tugas akhir ini dapat bermanfaat. Penulis menyadari sepenuhnya bahwa tugas akhir ini masih jauh dari sempurna, oleh karena itu segala kritik dan saran yang sifatnya membangun selalu penulis harapkan. Semoga tugas akhir ini dapat bermanfaat bagi penulis khususnya dan bagi semua yang membutuhkan umumnya. Akhir kata, semoga Allah SWT selalu melimpahkan rahmat serta hidayah-Nya kepada kita semua, Amin amin ya robbal 'alamiin

Wassalamualaikum Wr.Wb

Yogyakarta, 14 Maret 2018

Penulis

Riza Indriani Rakhmalia

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN PEMBIMBING	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERSEMBAHAN	iv
KATA PENGANTAR	v
DAFTAR ISI	vi
DAFTAR TABEL	viii
DAFTAR GAMBAR	ix
DAFTAR LAMPIRAN	x
DAFTAR ISTILAH/TAKARIR	xi
PERNYATAAN	xii
INTISARI	xiii
ABSTRACT	xiv
BAB I. PENDAHULUAN	1
1.1. Latar Belakang Masalah	1
1.2. Rumusan Masalah	6
1.3. Batasan Masalah	6
1.4. Tujuan Penelitian	6
1.5. Manfaat Penelitian	7
BAB II. TINJAUAN PUSTAKA	8
BAB III. LANDASAN TEORI	12
3.1. Curah Hujan	12
3.1.1. Geografis Jawa Timur	12
3.1.2. Variabel Berpengaruh dalam Curah Hujan	12
3.2. Data <i>Mining</i>	13
3.2.1. <i>Assosiation Rule Mining</i>	13
3.2.2. <i>Clustering</i>	14
3.2.3. Klasifikasi	14
3.3. Tahap Klasifikasi	14

3.3.1.	<i>Data Collection</i>	14
3.3.2.	<i>Data Cleaning</i>	14
3.3.3.	<i>Data Reduction</i>	14
3.4.	<i>Data Training</i> dan <i>Testing</i>	15
3.5.	<i>Naïve Bayes Classifier</i>	15
3.6.	<i>Support Vector Machine</i>	17
3.7.	<i>Non-Linear Separable Data</i>	17
3.8.	<i>Confussion Matrix</i>	21
BAB IV.	METODOLOGI PENELITIAN	23
4.1.	Populasi dan Sampel	23
4.2.	Sumber Data	23
4.3.	Variabel Penelitian.....	23
4.3.1.	Variabel Dependen.....	23
4.3.2.	Variabel Independen	24
4.4.	Metode Pengumpulan Data.....	24
4.5.	Metode Analisis Data	25
4.6.	Tahapan Analisis	25
BAB V.	HASIL DAN PEMBAHASAN	28
5.1.	Analisis deskriptif.....	28
5.2.	Klasifikasi dengan NBC dan SVM	33
5.2.1.	Pembagian Data <i>Training</i> dan Data <i>Testing</i>	33
5.2.2.	Analisis Naive Bayes Classifier.....	33
5.2.3.	Analisis Klasifikasi <i>Support Vector Machine</i>	35
5.3.	Perbandingan Hasil Metode NBC dan SVM	39
BAB VI.	PENUTUP	42
6.1.	Kesimpulan	42
6.2.	Saran	43
DAFTAR PUSTAKA	44
Lampiran	47

DAFTAR TABEL

Tabel 3.1. Tabel <i>Confusion Matrix</i>	21
Tabel 4.1. Variabel Penelitian	23
Tabel 4.2. Kategori Variabel Dependen	24
Tabel 4.3. Variabel Independen.....	24
Tabel 5.1. Penggunaan Data <i>Training</i> dan Data <i>Testing</i>	33
Tabel 5.2. <i>Priori Probability</i>	34
Tabel 5.3. Hasil <i>Confusion Matrix</i> NBC Data <i>Training</i>	34
Tabel 5.4. Akurasi SVM dengan Kernel RBF Data <i>Training</i>	36
Tabel 5.5. Akurasi SVM dengan Kernel RBF Data <i>Testing</i>	36
Tabel 5.6. <i>Confusion Matrix</i> SVM dengan Kernel RBF (C=1, Gamma=1)	37
Tabel 5.7. Akurasi SVM dengan Kernel <i>Polynomial</i> Data <i>Training</i>	38
Tabel 5.8. Akurasi SVM dengan Kernel <i>Polynomial</i> Data <i>Testing</i>	38
Tabel 5.9. <i>Confussion Matrix</i> SVM dengan Kernel Polynomial (C=5, Degree=1) .	39
Tabel 5.10. Hasil Perbandingan.....	39
Tabel 5.11. <i>Confusion Matrix</i> Data <i>Testing</i> SVM RBF	40

DAFTAR GAMBAR

Gambar 1.1. Curah Hujan Tahun 2013 - 2017	3
Gambar 1.2. Jumlah Banjir di Indonesia Tahun 2013 - 2017.....	4
Gambar 1.3. Jumlah Banjir Terbanyak di Indonesia Tahun 2013 – 2017	5
Gambar 3.1. <i>Hyperplane</i>	18
Gambar 4.1. <i>Flowchart</i> Penelitian.....	25
Gambar 5.1. Jumlah Status Hujan Jawa Timur Tahun 2013 - 2017.....	29
Gambar 5.2. Jumlah Status Hujan Ekstrem Jawa Timur Tahun 2013 - 2017	29
Gambar 5.3. Jumlah Status Tidak Hujan Jawa Timur Tahun 2013 - 2017	30
Gambar 5.4. Deskriptif Variabel Lama Penyinaran dan Kecepatan Angin.....	31
Gambar 5.5. Deskriptif Variabel Kelembaban dan Suhu	32

DAFTAR LAMPIRAN

Lampiran 1. Data Penelitian	48
Lampiran 2. <i>Syntax NBC</i>	49
Lampiran 3. <i>Output NBC</i>	51
Lampiran 4. <i>Syntax SVM</i>	55
Lampiran 5. <i>Output SVM dengan Kernel RBF</i>	57
Lampiran 6. <i>Output SVM dengan Kernel Polynomial</i>	59

DAFTAR ISTILAH/TAKARIR

ANN	: <i>Artificial Neural Network</i>
BMKG	: Badan Meteorologi dan Klimatologi Geofisika
NBC	: <i>Naïve Bayes Classifier</i>
POT	: <i>Peals Over Threshold</i>
SVM	: <i>Support Vector Machine</i>
RBF	: <i>Radial Basis Function</i>

PERBANDINGAN METODE PERNYATAAN
SUPPORT VECTOR MACHINE DAN NAIVE BAYES CLASSIFIER DALAM
PREDIKSI CURAH HUJAN

Dengan ini menyatakan bahwa dalam skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan disuatu Perguruan Tinggi dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain kecuali yang secara tertulis diacu dalam naskah ini dan diterbitkan dalam daftar pustaka.

Universitas Islam Indonesia

Yogyakarta, 15 April 2018

Penulis



Riza Indriani Rakhmalia

INTISARI

Curah hujan merupakan salah satu elemen pada iklim. Intensitas curah hujan yang tinggi atau yang sering disebut hujan ekstrem dapat mengakibatkan kerusakan besar. Untuk meminimalisir kerusakan bencana yang diakibatkan oleh curah hujan maka diperlukan prediksi curah hujan harian. Prediksi curah hujan telah menjadi salah satu masalah yang paling utama dan teknologi mendukung seluruh dunia pada saat terakhir. Metode Naive Bayes Classifier (NBC) dan Support Vector Machine (SVM) merupakan metode klasifikasi yang dapat digunakan untuk prediksi curah hujan pada penelitian ini. Provinsi Jawa Tengah merupakan provinsi dengan curah hujan yang tinggi. Oleh karena itu pada penelitian ini akan membahas mengenai klasifikasi curah hujan dengan metode NBC dan SVM. Data yang digunakan berasal dari BMKG dari tahun 2013 hingga tahun 2017 dengan empat variabel independen. Hasil dari penelitian menunjukkan untuk data training metode NBC menghasilkan akurasi sebesar 0,784935, metode SVM dengan kernel RBF-Cat dan Gamma=1 akurasi sebesar 0,805916 dan metode SVM dengan kernel polinomial C=3 dan Degree=1 akurasi sebesar 0,79011. Oleh karena itu model yang digunakan untuk prediksi curah hujan harian berikutnya adalah metode SVM dengan Kernel RBF-Cat, Gamma=1.

Kata Kunci : Curah Hujan, Klasifikasi, Kernel, NBC, SVM

PERBANDINGAN HASIL METODE NAÏVE BAYES CLASSIFIER DAN SUPPORT VECTOR MACHINE DALAM KLASIFIKASI CURAH HUJAN

(Studi Kasus : Curah Hujan di Jawa Timur Tahun 2013 - 2017)

Riza Indriani Rakhmalia

Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Islam Indonesia

INTISARI

Curah hujan merupakan salah satu elemen pada iklim. Intensitas curah hujan yang tinggi atau yang sering disebut hujan ekstrem dapat mengakibatkan terjadinya banjir. Untuk meminimalisir terjadinya banjir yang diakibatkan oleh curah hujan, maka diperlukan prakiraan curah hujan harian. Prakiraan curah hujan telah menjadi salah satu masalah yang paling ilmiah dan teknologi menantang seluruh dunia pada abad terakhir. Metode Naïve Bayes Classifier (NBC) dan Support Vector Machine (SVM) merupakan metode klasifikasi yang dapat digunakan untuk prakiraan curah hujan pada penelitian ini. Provinsi Jawa Timur merupakan provinsi dengan curah hujan yang tinggi. Oleh karena itu, pada penelitian ini akan membahas mengenai klasifikasi curah hujan dengan metode NBC dan SVM. Data yang digunakan berasal dari BMKG dari tahun 2013 hingga tahun 2017 dengan empat variabel independen. Hasil dari penelitian menunjukkan untuk data training metode NBC menghasilkan akurasi sebesar 0,7830805, metode SVM dengan kernel RBF C=1 dan Gamma=1 akurasi sebesar 0,806018 dan metode SVM dengan kernel polynomial C=5 dan Degree=1 akurasi sebesar 0,79011. Oleh karena itu model yang digunakan untuk prediksi curah hujan harian berikutnya adalah metode SVM dengan Kernel RBF (C=1, Gamma=1).

Kata Kunci : Curah Hujan, Klasifikasi, Kernel, NBC, SVM.

**THE COMPARISON RESULT OF NAÏVE BAYES CLASSIFIER AND
SUPPORT VECTOR MACHINE METHOD IN RAINFALL
CLASSIFICATION**

(Case Study : East Java Rainfall in 2013 to 2017)

Riza Indriani Rakhmalia

Department Statistics, Faculty of Mathematics and Natural Science
Islamic University of Indonesia

ABSTRACT

Rainfall is one of element in the climate. The intensity of high rainfall or called extreme rain can lead to flooding. To minimize the occurrence of floods caused by rainfall, the daily rainfall forecast is required. Weather forecasting has been one of the most scientifically and technologically challenging problems around the world in the last century. Naïve Bayes Classifier (NBC) and Support Vector Machine (SVM) methode is classification method that can be used to rainfall forecast in this research. East Java is the province with the high of rainfall. Therefore, in this research will be discuss about classification of rainfall using NBC and SVM method. This research used data in BMKG from 2013 until 2017 with four independent variable. The result of this research shown for training data NBC method has the accuracy 0,7830805, SVM method with RBF kernel $C=1$ and $\Gamma=1$ has accuracy 0,806018, and for SVM method with polynomial kernel $C=5$ and $Degree=1$ has accuracy 0,79011. Therefore, the method can used for predicting the daily rainfall in the next period is SVM method with RBF kernel ($C=1$, $\Gamma=1$).

Key Word :Classification, Kernel, NBC, Rainfall, SVM

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Data Mining merupakan kumpulan dari kegiatan yang meliputi pengumpulan dan pemakaian data masa lalu untuk menemukan pola atau hubungan dalam data yang berukuran besar. *Output* pada data *mining* tersebut dapat dijadikan pengambilan keputusan dimasa depan. Metode ini merupakan gabungan dari 4 disiplin ilmu yaitu visualisasi, statistik, basis data dan *machine learning*. Adapun klasifikasi merupakan salah satu ilmu yang terdapat pada *machine learning*. Klasifikasi merupakan salah satu metode yang dapat menangani *big data*. Terdapat banyak metode yang ada dalam klasifikasi diantaranya adalah *Naïve Bayes Classifier*, *Support Vector Machine*, *Artificial Neural Network*, *Classification Tree*, *K-Nearest Neighbour*, dan lain-lain. Pada penelitian ini metode klasifikasi yang digunakan adalah metode perbandingan antara *Naïve Bayes Classifier*, dan *Support Vector Machine*.

Naïve Bayes Classifier (NBC) atau yang sering disebut *Bayessian Classification* merupakan metode yang menangani masalah dengan mencari nilai peluang. Metode NBC ini salah satu metode klasifikasi yang efektif, mudah, efisien dan handal dalam mengatasi data seperti atribut yang kurang atau hilang. *Dataset* yang besar baik dalam bentuk diskrit maupun kontinu juga dapat menggunakan *Naïve Bayes*. (Subhan, dan Zainul, 2015). Mudahnya metode NBC membuat metode ini banyak digunakan dalam penelitian. Selain metode NBC tersebut, terdapat metode yang dapat juga digunakan untuk data kontinu, salah satunya adalah SVM.

Metode *Support Vector Machine* (SVM) mempunyai kelebihan yaitu kemampuan dalam generalisasi atau mampu mengklasifikasikan data yang tidak termasuk data *training* secara tepat, tingkat generalisasi tidak dipengaruhi oleh dimensi peubah yang diamati dan proses komputasi yang lebih cepat karena hanya

melibatkan beberapa observasi saja dalam pembentukan fungsi keputusannya (Siregar, 2017). Pada dasarnya untuk mengetahui metode mana yang lebih bagus, maka perlu dicari nilai akurasi pada masing-masing metode. Nilai akurasi yang tinggi yang nantinya digunakan untuk melakukan prediksi. Metode klasifikasi NBC maupun SVM dapat diaplikasikan diberbagai bidang, salah satunya adalah bidang meteorologi dan klimatologi. Salah satu aplikasi dalam bidang meteorologi dan klimatologi adalah dalam hal prakiraan hal iklim dan curah hujan.

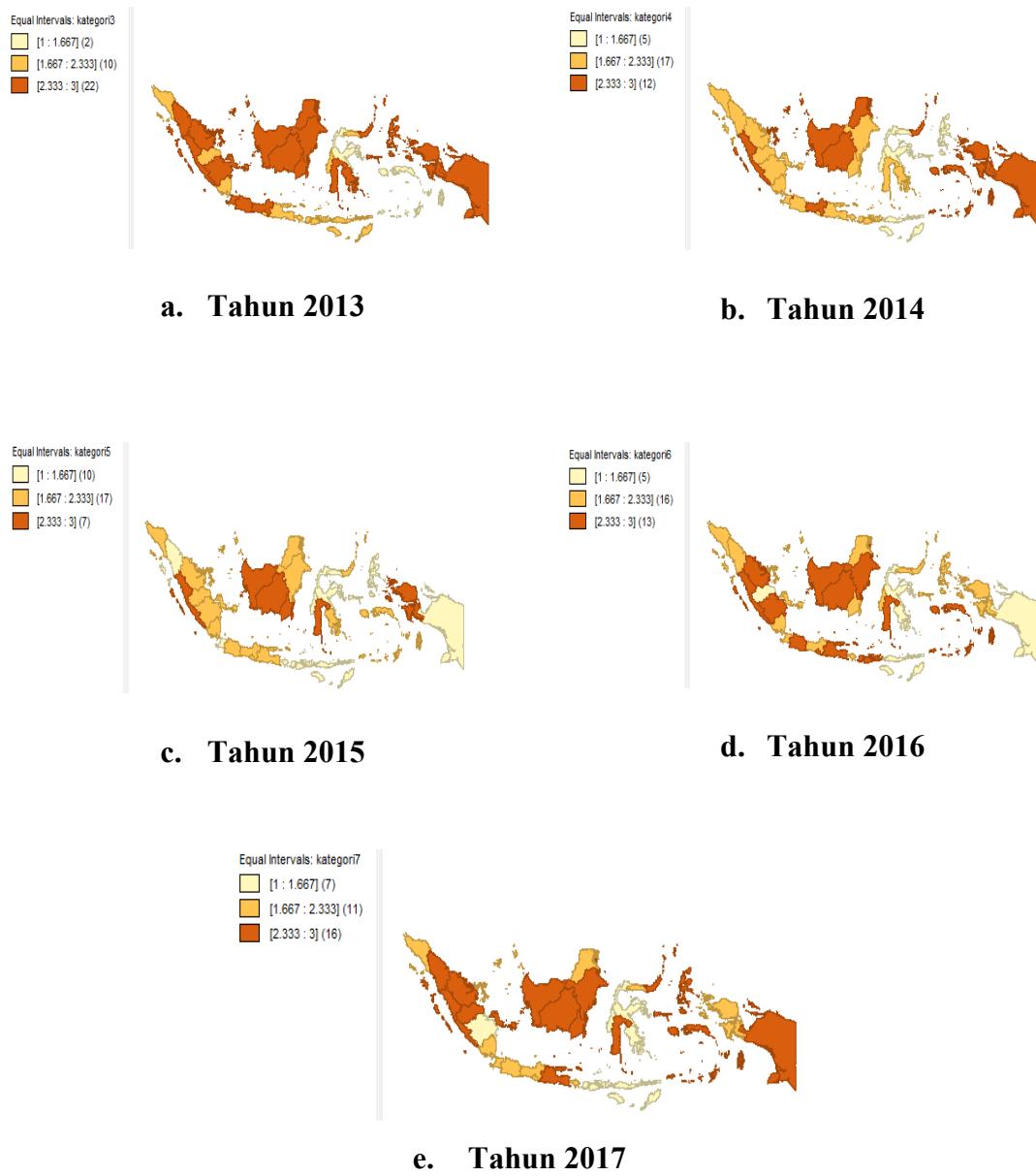
Iklim merupakan unsur geografi yang penting, karena iklim memiliki pengaruh yang besar terhadap aktivitas manusia di lingkungan, seperti bidang transportasi, pertanian, perkebunan, dan lain-lain. Terdapat beberapa macam iklim di dunia. Iklim merupakan cuaca yang terjadi pada rentang waktu yang panjang dan wilayah yang luas. Cuaca merupakan kondisi udara pada wilayah tertentu dan dalam waktu yang relatif pendek. Cuaca dapat berubah dalam waktu singkat ditandai dengan adanya siang dan malam. Salah satu faktor penting dalam perubahan cuaca adalah curah hujan.

Curah hujan merupakan salah satu elemen pada iklim. Curah hujan juga merupakan faktor yang berpengaruh langsung terhadap perubahan cuaca, baik perubahan cuaca yang baik maupun perubahan cuaca yang buruk. Curah hujan adalah jumlah air yang jatuh ke permukaan tanah yang kemudian dihitung dengan menggunakan satuan dalam milimeter diatas permukaan tanah yang datar. (Muhammad, Fhira dan Adiwijaya, 2015). Terdapat beberapa faktor utama yang dapat mempengaruhi terjadinya curah hujan antara lain suhu udara, kelembaban udara, kecepatan angin, dan lama penyinaran.

Prakiraan curah hujan telah menjadi salah satu masalah yang paling ilmiah dan teknologi menantang seluruh dunia pada abad terakhir. Ini disebabkan oleh dua faktor utama yaitu digunakan untuk melakukan berbagai kegiatan manusia dan kemajuan komputasi. Untuk membuat prediksi yang akurat adalah salah satu tantangan utama yang dihadapi ahli Meteorologi seluruh dunia. Sejak zaman kuno, prediksi cuaca maupun curah hujan telah menjadi salah satu paling menarik.

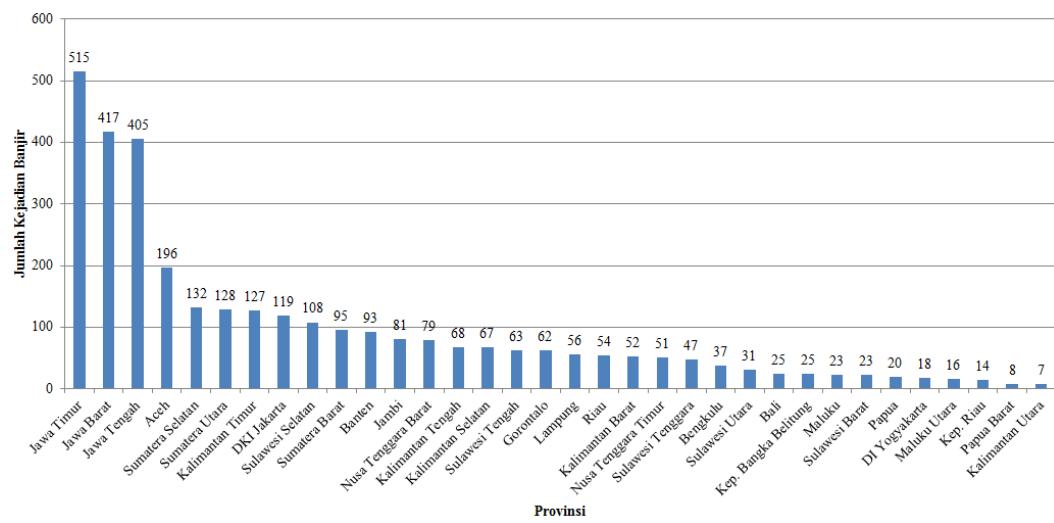
Para ilmuwan telah mencoba untuk meramalkan meteorologi karakteristik menggunakan sejumlah metode.

Indonesia memiliki iklim tropis, dengan iklim tersebut akibatnya penguapan air ke udara menjadi sangat besar, sehingga intensitas curah hujan menjadi tidak stabil. Intensitas curah hujan yang tinggi atau yang sering disebut hujan ekstrem dapat mengakibatkan terjadinya banjir. Hujan ekstrem ini terjadi jika curah hujan mencapai >50 mm/hari (BMKG).



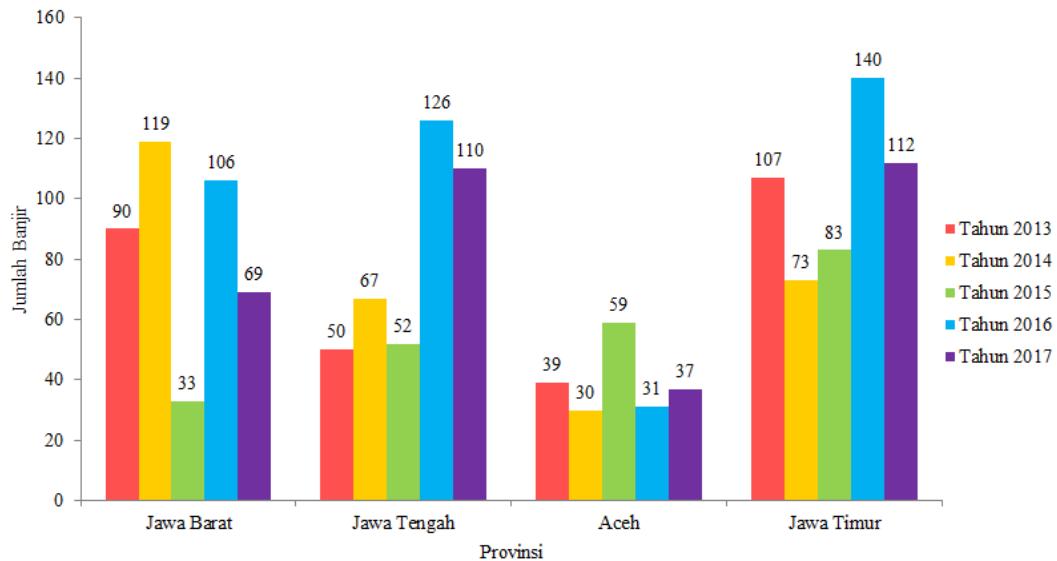
Gambar 1.1 Curah Hujan Tahun 2013 – 2017

Berdasarkan **gambar 1.1** merupakan peta kategori curah hujan di Indonesia dari tahun 2013 hingga tahun 2017. Terdapat tiga kategori dengan kategori pertama curah hujan ringan, kategori kedua curah hujan sedang, dan kategori ketiga curah hujan tinggi. Indonesia mengalami perubahan curah hujan dari tahun 2013 hingga tahun 2017, terdapat provinsi yang mengalami penurunan curah hujan, tetapi ada juga yang mengalami peningkatan curah hujan. Provinsi Jawa Timur mengalami peningkatan curah hujan pada tahun 2016 dan tahun 2017. Tingginya curah hujan dapat menyebabkan terjadinya banjir. Berikut merupakan jumlah kejadian banjir di Indonesia.



Gambar 1.2 Jumlah Banjir di Indonesia dari Tahun 2013 - 2017

Berdasarkan **gambar 1.2** diketahui bahwa untuk empat Provinsi dengan jumlah bencana banjir paling banyak dari tahun 2013 – 2017 adalah provinsi Aceh, Jawa Timur, Jawa Tengah, dan Jawa Barat. Menurut BBC news bencana yang paling banyak terjadi di Indonesia adalah banjir dengan jumlah kejadian 2.342 peristiwa banjir pada tahun 2016. Jumlah kejadian ini meningkat 35 % dari tahun 2015 (BBC news, 2016).



Gambar 1.3 Jumlah Banjir Terbanyak di Indonesia Tahun 2013-2017

Pada **gambar 1.3** Provinsi Jawa Timur merupakan salah satu Provinsi dengan jumlah bencana banjir paling banyak dari tahun 2013 hingga tahun 2017. Selain itu, Jawa Timur juga merupakan provinsi dengan padat penduduk. Banyaknya jumlah banjir di Jawa Timur diikuti dengan banyaknya penduduk. Oleh karena itu, pada penelitian ini peneliti memfokuskan pada Provinsi Jawa Timur. Berdasarkan **gambar 1.3** diperoleh bahwa untuk setiap tahunnya Jawa Timur merupakan provinsi yang paling banyak jumlah bencana banjir, kecuali pada tahun 2014.

Terdapat tiga status hujan pada penelitian ini yang ditentukan peneliti yaitu “Tidak hujan”, ”Hujan”, dan ”Hujan Ekstrem”, adanya status hujan ekstrem tersebut dikarenakan banyaknya bencana banjir yang diakibatkan oleh hujan ekstrem di Indonesia. Oleh karena itu, berdasarkan pernyataan diatas maka prediksi terhadap curah hujan diperlukan adanya penelitian prediksi curah hujan di Jawa Timur, guna mengantisipasi perubahan cuaca yang tidak menentu yang dapat mengakibatkan banjir di Jawa Timur.

1.1 Rumusan Masalah

Berdasarkan uraian dari latar belakang masalah diatas maka diperoleh rumusan masalah sebagai berikut:

1. Bagaimana kondisi secara umum curah hujan di Jawa Timur pada Tahun 2013 – 2017?
2. Bagaimana hasil implementasi antara metode *Naïve Bayes Classifier* dan *Support Vector Machine*?
3. Bagaimana hasil perbandingan antara metode *Naïve Bayes Classifier* dan *Support Vector Machine*?

1.2 Batasan Masalah

Adapun batasan masalah pada penelitian ini adalah sebagai berikut:

1. Data yang digunakan adalah data BMKG dan data BNPB dari tahun 2013 sampai tahun 2017 di Provinsi Jawa Timur.
2. Metode yang digunakan adalah metode *Naïve Bayes Classifier* dan *Support Vector Machine*
3. *Software* yang digunakan adalah R.3.3.2
4. Untuk variabel status hujan yang digunakan hanya Hujan, Hujan Ekstrem, dan Tidak Hujan.
5. Batasan untuk nilai akurasi yang digunakan adalah lebih besar 78% untuk data *training*.
6. Untuk semua variabel yang digunakan merupakan data hasil rata-rata yang diperoleh langsung dari website bmkg.go.id.

1.3 Tujuan Penelitian

Penelitian ini memiliki tujuan sebagai berikut:

1. Untuk mengetahui kondisi umum curah hujan di Jawa Timur pada Tahun 2013 – 2017?
2. Untuk mendapatkan hasil implementasi antara metode *Naïve Bayes Classifier* dan *Support Vector Machine*?
3. Untuk mendapatkan hasil perbandingan antara metode *Naïve Bayes Classifier* dan *Support Vector Machine*?

1.4 Manfaat Penelitian

Penelitian ini dibuat dengan memiliki manfaat, adapun manfaat penelitian ini adalah:

1. Dengan didapatkannya hasil dari data *training*, maka dapat diketahui seberapa baik hasil pembelajaran yang diberikan dari data, sehingga dapat digunakan untuk menguji pada data *testing*.
2. Dengan didapatkannya hasil dari data *testing* yang baik, maka dapat digunakan untuk melakukan prediksi kedepannya.
3. Dapat menjelaskan penggunaan metode NBC dan SVM.
4. Penelitian ini dapat memberikan informasi mengenai karakteristik status hujan, tidak hujan, dan hujan ekstrem yang kemudian dapat digunakan oleh BMKG Jawa Timur dalam mengantisipasi hujan ekstrem.

BAB II

TINJAUAN PUSTAKA

Setelah peneliti melakukan telaah terhadap beberapa penelitian, terdapat beberapa penelitian yang memiliki keterkaitan dengan penelitian yang akan peneliti lakukan. Baik keterkaitan pada metode maupun pada subyek penelitian. Berikut merupakan beberapa keterkaitan penelitian terdahulu.

Penelitian mengenai pemanfaatan data *mining* untuk prakiraan cuaca oleh Mujiasih (2011). Adapun tujuan dari penelitian ini adalah memperoleh model prakiraan yang sesuai agar dapat memudahkan dalam analisa dan prakiraan cuaca. Data yang digunakan adalah data sinoptik 9 Stasiun Maritim tahun 2009. Metode klasifikasi yang digunakan adalah *Association Rule*, *Classification Tree*, dan *Random forest*. Untuk metode *Association Rule* menggunakan seluruh variabel yaitu tanggal, stasiun, arah angin, kecepatan angin, suhu udara, suhu titik embun, tutupan awan, dan curah hujan. Untuk metode *Classification Tree*, dan *Random forest* menggunakan data suhu udara, suhu titik embun, tutupan awan, dan kecepatan angin. Data *training* 80 % dan untuk data *testing* sebesar 20 %. Hasil yang diperoleh adalah untuk metode *Association Rule* menghasilkan nilai akurasi sebesar 60,9 %, *Random forest* sebesar 64,6 %, *Classification Tree* sebesar 64,4 %, dan untuk C4.5 sebesar 69,5 %. Oleh karena itu model prediksi yang dipilih adalah metode C4.5.

Penelitian yang dilakukan oleh Olaiya pada tahun 2012. Pada penelitian ini membahas mengenai aplikasi dari teknik data *mining* untuk memprediksi cuaca dan perubahan iklim. Variabel yang digunakan adalah suhu maksimum, curah hujan, evaporasi, dan kecepatan angin. Teknik data *mining* yang digunakan adalah ANN dan *Decision Tree*. Data yang digunakan pada penelitian ini bersumber dari Badan Meterologi Ibadan, Nigeria dari tahun 2000 sampai tahun 2009. Metode yang digunakan adalah *Decision Tree* dengan *error* yang dihasilkan mencapai 33,3% untuk data *training* dan 25,0% untuk data *testing*, selanjutnya metode yang

digunakan adalah ANN dengan *error* yang dihasilkan untuk data *training* sebesar 19,8% dan untuk data *testing* sebesar 28,2%.

Penelitian berikutnya juga dilakukan oleh Octaviani, dkk (2014). Tujuan dari penelitian ini adalah untuk menerapkan metode SVM pada klasifikasi akreditasi Sekolah Dasar (SD) di Magelang. Dengan menggunakan data yang diperoleh dari *website* resmi BAN-S/M dengan 80 % data *training* dan 20 % data *testing*. Variabel dependen terdiri dari tiga kategori yaitu akreditasi sekolah, sedangkan untuk variabel independen terdapat tujuh variabel yaitu antara lain standar isi, standar proses, standar kompetensi lulusan, standar pendidik dan kependidikan, standar sarana dan prasarana, standar pengelolaan, dan standar penilaian pendidikan. Hasil dari penelitian Puspita ini adalah dengan menggunakan metode SVM nilai akurasi untuk kernel RBF lebih besar yaitu 100 % dari data *training*, dan 98.81 % dari data *testing* dengan nilai sigma 3 dan C 1. Sedangkan, untuk kernel Polynomial akurasi yang didapat untuk data *training* sebesar 93.902 % dan untuk data *testing* sebesar 92.683 %.

Pada penelitian selanjutnya dilakukan oleh Assaffat (2015). Adapun tujuan dari penelitian ini adalah untuk menganalisis akurasi peramalan beban listrik harian pada sektor industri menggunakan metode SVM dengan kernel RBF. Data pada penelitian Luqman ini didapat dari salah satu industri farmasi yaitu PT.Phapros Indonesia yaitu data beban listrik harian tahun 2014. Variabel yang digunakan adalah beban listrik harian, kapasitas produksi dan jenis haria kerja. Hasil dari penelitian ini diperoleh nilai *errornya* sebesar 2.63 %.

Selanjutnya, penelitian yang dilakukan oleh Subhan dan Ahmad (2017). Tujuan dari penelitian ini adalah dapat mempermudah proses dalam penentuan cuaca dengan nilai akurasi yang baik. Metode yang digunakan adalah Naive Bayes Calssifier dengan data curah hujan harian. Klasifikasi yang dilakukan adalah klasifikasi hujan dan tidak hujan. Hasil yang diperoleh adalah nilai akurasi sebesar 82,5%, nilai recall sebesar 82,6%, dan nilai presisi sebesar 80%.

Penelitian berikutnya dilakukan oleh Novandy dan Oktria (2017). Penelitian ini bertujuan untuk memperoleh pola klasifikasi yang dapat digunakan untuk memprediksi cuaca kedepan. Metode yang digunakan pada penelitian ini adalah algoritma C4.5. Data diperoleh dari *World Weather Online* mulai tanggal 12 Agustus 2016 hingga 20 Agustus 2016. Variabel dalam penelitian ini antara lain *date*, *time*, *description*, *weather*, *temperature*, *rain*, *wind*, *direction* (arah angin), *cloud*, *humidity*, dan *pressure*. Hasil penelitian ini menunjukkan bahwa dari 10 variabel yang diteliti terdapat empat variabel yang berpengaruh dalam klasifikasi cuaca. Pada penelitian ini juga dibentuk sebuah *website* dari hasil yang sudah diperoleh. Untuk nilai akurasi pada penelitian ini sebesar 88.89 %.

Penelitian selanjutnya dilakukan oleh Siregar tahun 2017. Tujuan dari penelitian tersebut adalah membuat pemodelan SVM untuk klasifikasi curah hujan di Indramayu. Metode SVM dengan menggunakan *soft Margin hyperplane*, *kernel gaussian*, dan *kernel polynomial*. Dalam penelitian ini data yang digunakan adalah data sekunder yang berasal dari data penlitian Saleh (2015) mulai tahun 1981 hingga tahun 2014. Penelitian ini untuk kategorisasi dalam curah hujan dengan menggunakan kategori BMKG dan kategori dengan metode *Peals Over Threshold* (POT). Hasil penelitian ini menunjukkan bahwa algoritma SVM dengan menggunakan kernel gaussian lebih tinggi nilai akurasinya jika dibandingkan dengan kerjenl *polynomial*, dan *soft margin*. Sedangkan, berdasarkan batasan pengkategorian curah hujan, diperoleh nilai akurasi dengan metode SVM menggunakan kategori POT 9 % lebih tinggi, jika dibandingkan dengan kategori BMKG.

Penelitan tahun 2017 juga telah dilakukan oleh Fibrianda dan Adhitya Bhawiyuga mengenai perbandingan akurasi deteksi serangan pada jaringan komputer dengan menggunakan metode NBC dan SVM. Tujuan dari penelitian Mercury adalah untuk melihat tingkat akurasi pada metode NBC, SVM Linear, SVM *Polynomial*, dan SVM Sigmoid. Data yang digunakan pada penelitian ini adalah dataset dari ISCX2012 testbed tanggal 14 Juni 2012. Hasil yang diperoleh pada penelitian ini adalah akurasi untuk NBC, SVM Linear, SVM *Polynomial*,

dan SVM Sigmoid berturut turut adalah 85,055 %, 99,995 %, 99,999 %, dan 99,995 %. Oleh karena itu akurasi yang tertinggi adalah SVM Plynomial sedangkan untuk akurasi yang paling rendah adalah NBC.

Pada penelitian tugas akhir ini dengan penelitian sebelumnya terdapat perbedaan yaitu dimana pada penelitian tugas akhir ini menggunakan data curah hujan harian pada Provinsi Jawa Timur dan dianalisis dengan menggunakan dua metode yaitu matode NBC dan SVM. Selain itu, penelitian pada tugas akhir ini juga bertujuan untuk membandingkan antara metode NBC dan SVM terhadap kedua metode tersebut. Oleh karena itu mengingat, subjek, objek dan tempat penelitian yang berbeda, maka penulis tertarik untuk melakukan penelitian tentang Perbandingan Hasil Metode *Naïve Bayes Classifier* dan *Support Vector Machine* Dalam Klasifikasi Curah Hujan (Studi Kasus : Curah Hujan di Jawa Timur Tahun 2013 - 2017).

BAB III

LANDASAN TEORI

3.1 Curah Hujan

Curah hujan merupakan jumlah air yang jatuh pada permukaan tanah yang datar. Adapun satuan curah hujan selalu dinyatakan dalam milimeter (mm). Curah hujan 1 milimeter artinya tempat yang datar dengan luasan satu meter persegi dapat tertampung air setinggi satu milimeter atau tertampung air sebanyak satu liter (Prawaka, 2016). Curah hujan yang tinggi biasanya disebut sebagai musim hujan, sedangkan curah hujan yang rendah biasanya disebut dengan musim kemarau.

1.1.1. Geografis Jawa Timur

Jawa Timur merupakan salah satu provinsi di Indonesia. Jawa Timur yang memiliki ibukota Surabaya dengan jumlah penduduk menurut sensus 2015 ada 42.030.633 jiwa dengan luas wilayah 47.922 km^2 . Jawa Timur mempunyai iklim tropis, dengan rata-rata curah hujan tahunan 1.900 meter. Suhu di Jawa Timur rata-rata berkisar 21°C hingga 34°C .

1.1.2. Variabel Berpengaruh dalam Curah Hujan

a. Rata-rata Suhu ($^\circ\text{C}$)

Menurut BMKG “Suhu udara merupakan ukuran energi kinetik rata-rata dari pergerakan molekul. Alat pengukur suhu pada umumnya yang digunakan pada BMKG adalah thermometer kaca (*liquid-in-glass thermometer*) untuk peralatan Konvensional dan *thermometer* PT-100 untuk peralatan-peralatan digital”. Data untuk variabel rata-rata suhu adalah numerik dengan satuan ($^\circ\text{C}$).

b. Rata-rata kelembaban (%)

Kelembaban udara/ legas udara menurut BMKG adalah “jumlah kandungan uap air yang ada dalam udara. Kandungan uap air di udara berubah-ubah bergantung pada suhu makin tinggi suhu, makin banyak kandungan uap airnya. Alat pengukur kelembaban udara adalah higrometer.” Pada variabel independen

rata-rata kelembaban ini data yang digunakan berskala numerik dalam bentuk persentase.

a. Lama Penyinaran (jam)

Lama penyinaran artinya adalah lama penyinaran matahari. Menurut BMKG “lama penyinaran merupakan salah satu dari beberapa unsur klimatologi, dan didefinisikan sebagai kekuatan matahari yang melebihi 120 W/m^2 . Alat pengukur penyinaran matahari adalah *Campbell Stokes Recorder*, alat ini secara resmi yang digunakan oleh BMKG”. Data pada variabel lama penyinaran dalam satuan jam.

b. Rata-rata kecepatan (knot)

Kecepatan angin rata-rata adalah pergerakan udara per satuan waktu yang dinyatakan dalam satuan meter per detik (m/d), kilometer per jam (km/j), dan mil per jam (mi/j) atau knot. *Anenometer* merupakan alat untuk mengukur kecepatan angin yang biasa digunakan dalam BMKG. Data kecepatan angin rata-rata adalah data numerik dalam satuan knot. Knot merupakan satuan kecepatan dimana nilainya sama dengan 1 mil laut per jam atau sekitar 1,151 mil per jam.

3.1 Data Mining

Data *mining* merupakan proses pengaplikasikan dalam mengekstrak data. Data *mining* juga merupakan analisis sekumpulan data dalam menemukan hubungan tidak terduga serta untuk meringkas data dengan cara baru agar dapat mudah dipahami (Novandy, 2017).

Tiga teknik yang sering digunakan dalam data *mining* (Haryati dkk, 2015 dalam Estoatnowo, 2016) .

3.1.1 Association Rule Mining

Teknik dalam data *mining* untuk menentukan hubungan antara kombinasi atribut adalah *Association Rule Mining*. Contoh yang biasa digunakan adalah analisa dalam pembelian barang disuatu swalayan sehingga dapat menghasilkan strategi dalam penempatan barang di swalayan atau dengan menggunakan diskon untuk kombinasi barang tertentu.

3.1.2 Clustering

Clustering digunakan untuk memberikan label pada kelas data yang belum diketahui sehingga sering disebut *unsupervised learning*. Prinsip *clustering* adalah memaksimalkan kesamaan antar *cluster*.

3.1.3 Klasifikasi

Berbeda dengan clustering, klasifikasi digunakan untuk membedakan kelas yang telah diketahui, dengan tujuan dapat memprediksi kelas yang belum diketahui.

3.2 Tahap Klasifikasi

Klasifikasi suatu proses dalam mencari dan menentukan model atau fungsi yang dapat menjelaskan serta membedakan kelas data dengan tujuan adapat menggunakan data tersebut untuk memperkirakan kelas suatu objek yang statusnya tidak diketahui. Selama proses pembelajaran dalam pembuatan model, diperlukan suatu algoritma pembelajaran, antara lain yaitu SVM, *Naïve Bayes* , KNN, *Decision Tree*, ANN, dan lainnya.

3.2.1 Data Collection

Pada tahap ini dilakukan pengumpulan data yang digunakan dalam penelitian. Data yang terkumpul nantinya akan melalui tahap-tahap sebelum dilakukan analisis.

3.2.2 Data Cleaning

Data yang baik dan berkualitas adalah kunci dasar untuk menghasilkan data yang berkualitas, data *nois* data yang masih *outliers* atau *error*, data *incomplete* data yang nilai atributnya hilang, dan data *inconsistent* data yang tidak konsisten di dalam pengisian atributnya. Terdapat beberapa tahap pada data *cleaning*, yaitu (Ahmad Subhan, 2017)

- a. Identifikasi atau menghilangkan data *outlier*, dan menghilangkan data nois.
- b. Melengkapi data yang tidak lengkap, metode NBC dapat menangani data yang tidak lengkap.

3.2.3 Data Reduction

Pada tahap data *reduction* adalah tahap dimana data yang telah terkumpul, kemudian di bersihkan, proses selanjutnya adalah memilih variabel atau atribut

yang akan digunakan dalam penelitian. Tahap ini dilakukan untuk mengurangi atribut yang tidak digunakan akan tetapi tetap bersifat informatif (Ahmad Subhan, 2015).

3.3 Data Training dan Data Testing

Data umumnya dibagi menjadi data *training* dan data *testing*. Data *training* digunakan oleh algoritma klasifikasi (misalnya *Decision Tree*, Bayesian, *Neural Network*, SVM) untuk membentuk sebuah model klasifikasi. Model ini merupakan representasi pengetahuan yang akan digunakan untuk mengukur sejauh mana klasifikasi berhasil melakukan klasifikasi dengan benar. Karena itu, data yang ada yang ada pada data *testing* seharusnya tidak boleh ada pada data *training* sehingga dapat diketahui apakah model klasifikasi dapat melakukan klasifikasi dengan baik. Proporsi antara data *training* dan data *testing* tidak mengikat tetapi agar variasi dalam model tidak terlalu besar maka disarankan data *training* lebih besar dibandingkan data *testing*. Biasanya 2/3 dari total data dijadikan data *training* sedangkan sisanya dijadikan data *testing*. Selain itu, ada pula penelitian yang menghasilkan keakuratan model klasifikasi optimum dengan proporsi 80:20 dan 90:10 untuk data *training* dan data *testing* (Paratu, 2003 dalam Estoatnowo, 2016).

3.4 Naïve Bayes Classifier

Naïve Bayes Classifier (NBC) atau yang sering disebut juga *Bayesian Classification* merupakan proses metode klasifikasi yang digunakan untuk menentukan probabilitas suatu anggota dari suatu kelas. NBC merupakan teknik klasifikasi yang berbasis pada probablistik sederhana. Teorema bayes mempunyai kemampuan klasifikasi yang serupa dengan metode *Decision Tree* maupun *Neural Network*. Oleh karena itu, NBC juga efektif, efisien, dan handal dalam menangani *dataset* yang berukuran besar serta dapat menangani data yang tidak relevan. Simbol untuk X adalah vektor masukan yang berisi data dan Y adalah label kelas.

Rumus NBC untuk klasifikasi adalah sebagai berikut:

$$P(Y|X) = \frac{P(Y) \prod_{i=1}^q P(X_i|Y)}{P(X_i)} \quad \dots (3.1)$$

Setiap $X = \{X_1, X_2, X_3, \dots, X_q\}$ sebanyak q atribut atau q dimensi

dimana

$P(Y X)$	= Probabilitas data dengan vektor X pada kelas Y
$P(Y)$	= Probabilitas awal kelas Y (<i>Prior Probability</i>)
$P(X Y)$	= Probabilitas akhir (<i>Posterior Probability</i>)
$\prod_{i=1}^q P(X_i Y)$	= Probabilitas independen dari kelas Y dalam vektor X
i	= Observasi ke- i

Untuk nilai $P(X)$ akan selalu tetap sehingga yang dihitung hanya $P(Y) \prod_{i=1}^q P(X_i|Y)$ dengan memilih probabilitas terbesar yang nantinya akan digunakan sebagai kelas yang dipilih pada hasil prediksi, ini merupakan sistem kerja NBC.

NBC dapat digunakan untuk data kategorik maupun numerik. Untuk data numerik terdapat perlakuan khusus. Perlakuan khusus tersebut yaitu dengan mengasumsikan bentuk tertentu dari distribusi probabilistik dan mengestimasi parameter distribusi dengan data *training*. Distribusi Gaussian sering kali dipilih dalam mempresentasikan probabilitas bersyarat dari data kontinu pada kelas $P(X_i, Y)$. Terdapat dua parameter dalam Distribusi Gaussian yaitu *mean*, dan varian (Subhan, 2017). Dimana probabilitas bersyarat kelas Y_j untuk data X_i adalah:

$$P(X_i = x_i | Y = y_j) = \frac{1}{\sqrt{2\pi\sigma_{ij}^2}} \exp^{-\frac{(x_i - \mu_{ij})^2}{2\sigma_{ij}^2}} \quad \dots (3.2)$$

Dengan $i \geq \infty$ dan $j \geq \infty$

dimana:

X_i	= Atribut ke i
x_i	= Nilai atribut ke i
Y	= Kelas yang dicari
y_j	= Sub kelas Y yang dicari
μ_{ij}	= <i>Mean</i> sampel dari data <i>training</i> yang menjadi milik y_j
σ_{ij}^2	= Varian sampel data <i>training</i>

Terdapat beberapa karakteristik yang dimiliki oleh klasifikasi metode NBC, diantaranya adalah :

- a. Metode NBC ini *robust* terhadap data yang hilang maupun *outlier*.
- b. Tangguh terhadap atribut yang tidak relevan .

3.5 Support Vector Machine (SVM)

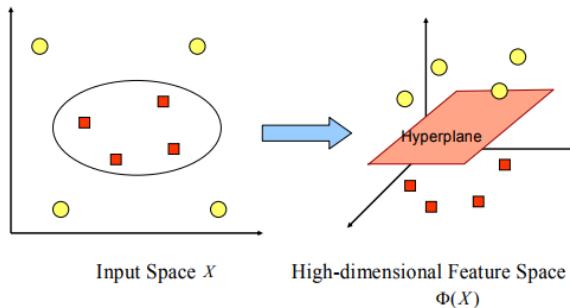
Pada tahun 1992 untuk pertama kali SVM diperkenalkan oleh Vapnik sebagai rangkaian konsep unggulan pada bidang *pattern recognition*. Usia SVM sebagai salah satu metode *pattern recognition* masih terbilang relatif muda. Dewasa ini SVM meruapakan salah satu metode yang berkembang pesat. SVM merupakan salah satu metode *machine learning* yang bekerja atas prinsip *Structural Risk Minimization* (SRM) yang bertujuan untuk menemukan *hyperplane* terbaik yang dapat memisahkan kelas-kelas pada *input space*.

SVM merupakan salah satu metode klasifikasi dalam data *mining*. SVM juga dapat melakukan prediksi baik pada klasifikasi maupun regresi (Santosa, 2007 dalam Pusphita, 2014). Pada dasarnya SVM memiliki prinsip linear, akan tetapi kini SVM telah berkembang sehingga dapat bekerja pada masalah *non-linear*. Cara kerja SVM pada masalah *non-linear* adalah dengan memasukkan konsep kernel pada ruang berdimensi tinggi. Pada ruang yang berdimensi ini, nantinya akan dicari pemisah atau yang sering disebut *hyperplane*. *Hyperplane* dapat memaksimalkan jarak atau *margin* antara kelas data. *Hyperplane* terbaik antara kedua kelas dapat ditemukan dengan mengukur *margin* dan kemudian mencari titik maksimalnya. Usaha dalam mencari *hyperplane* yang terbaik sebagai pemisah kelas-kelas adalah inti dari proses pada metode SVM (Cristianai, 2000 dalam Assaffat, 2015).

3.6 Non-Linear Separable Data

Dalam dunia nyata (*real world problem*) pada umumnya masalah data yang diperoleh jarang yang bersifat linear. Banyak yang bersifat *non linear*. Pada SVM sendiri terdapat fungsi yang disebut kernel. Fungsi kernel inilah yang digunakan untuk menyelesaikan problem *non linear*. Kernel berfungsi memungkinkan untuk

mengimplementasikan suatu model pada ruang dimensi lebih tinggi (ruang fitur).
(Nugroho, 2003)



Gambar 3.1 Hyperplane (Sumber : Nugroho, 2003)

Pada kasus SVM *non linear*, fungsi $\Phi(\vec{x})$ memetakan data \vec{x} ke ruang vektor yang berdimensi yang lebih tinggi. Pada ruang vektor yang baru ini, *hyperplane* yang memisahkan kedua class tersebut dapat dikonstruksikan. Hal ini sejalan dengan teori Cover yang menyatakan “*Jika suatu transformasi bersifat non linear dan dimensi dari feature space cukup tinggi, maka data pada input space dapat dipetakan ke feature space yang baru, dimana pattern-pattern tersebut pada probabilitas tinggi dapat dipisahkan secara linear*”.

Ilustrasi dari konsep ini dapat dilihat pada **gambar 3.2**. Awal mula diperlihatkan data berada pada *input space* berdimensi dua tidak dapat dipisahkan secara linear, selanjutnya terlihat bahwa fungsi Φ memetakan tiap data pada *input space* tersebut ke ruang vektor baru yang berdimensi lebih tinggi (dimensi 3), dimana kedua *class* dapat dipisahkan secara linear oleh sebuah *hyperplane*. Notasi matematika dari *mapping* ini adalah sebagai berikut:

$$\Phi : \mathbb{R}^d \rightarrow \mathbb{R}^q \quad d < q \quad \dots (3.3)$$

Pemetaan ini dilakukan dengan tujuan dua data yang berjarak dekat pada *input space* akan berjarak dekat juga pada *feature space*, sebaliknya jika dua data yang berjarak jauh pada *input space*, maka akan berjarak jauh juga pada *feature space*. Selanjutnya untuk proses pembelajaran SVM dalam menemukan *support vector*, hanya bergantung pada *dot product* berdasarkan data yang sudah ditransformasikan pada ruang baru yang berdimensi lebih tinggi, yaitu

$$\Phi(\vec{x}_i) \cdot \Phi(\vec{x}_j) \quad \dots (3.4)$$

Perhitungan *dot product* tersebut sesuai teori Mercer dapat digantikan dengan fungsi kernel $K(\vec{x}_i, \vec{x}_j)$ yang dapat mendefinisikan secara implisit transformasi Φ . Hal ini disebut sebagai *Kernel Trick*, yang dirumuskan

$$K(\vec{x}_i, \vec{x}_j) = \Phi(\vec{x}_i) \cdot \Phi(\vec{x}_j) \quad \dots (3.5)$$

Selanjutnya hasil klasifikasi dari data diperoleh dari persamaan berikut :

$$\begin{aligned} f(\Phi(\vec{x})) &= \vec{w} \cdot \Phi(\vec{x}) + b \\ &= \sum_{i=1}^n \alpha_i y_i \Phi(\vec{x}) \Phi(\vec{x}_i) + b \\ &= \sum_{i=1}^n \alpha_i y_i K(\vec{x}_i, \vec{x}) b \end{aligned} \quad \dots (3.6)$$

Dimana

- x_i = Data *input* baris ke-i
- x_j = Data *input* kolom ke-j
- y_i = Kelas *output* baris ke-i
- b = Bias
- α_i = *Support vector*
- w = Pembobot
- I q = q dimensi

Berikut ini merupakan beberapa fungsi kernel pada umumnya.

- a. Kernel Polynomial dengan Variabel Bebas q

$$K(\vec{X}_i, \vec{X}_j) = (\vec{X}_i \cdot \vec{X}_j + 1)^q \quad \dots (3.7)$$

- b. Kernel Gaussian atau RBF

$$K(\vec{X}_i, \vec{X}_j) = \exp\left(\frac{\|\vec{X}_i - \vec{X}_j\|^2}{2\sigma^2}\right) \quad \dots (3.8)$$

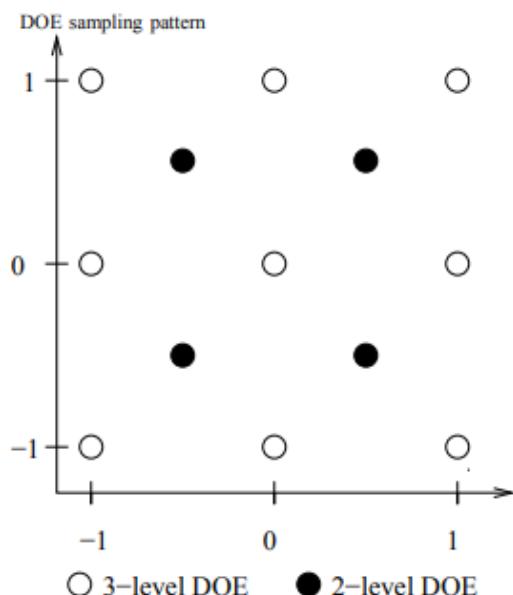
Algoritma SVM untuk klasifikasi (Santosa, 2007).

1. Hitung matriks kernel dengan menggunakan persamaan kernel, seperti pada persamaan 3.7 atau persamaan 3.8.
2. Tentukan pembatas untuk program kuadratik.
3. Tentukan fungsi tujuan seperti pada persamaan 3.6.

4. Selesaikan masalah program kuadratik.

3.7 Proses Tuning

SVM merupakan metode yang kuat untuk klasifikasi, akan tetapi pada berbagai rumus SVM peneliti diminta untuk menentukan dua atau lebih parameter dalam proses *training*, dan perubahan nilai pada parameter berpengaruh terhadap klasifikasi yang dihasilkan. Terdapat algoritma dengan prinsip *Design Of Experiment* (DOE) yang dapat mengoptimalkan nilai parameter lebih cepat. Algoritma ini telah berhasil di terapkan untuk beberapa aplikasi didunia nyata.



Gambar 3.3 Proses DOE (Sumber : Nugroho, 2003)

Pada **gambar 3.3** menunjukkan pola untuk pencarian dua parameter, yang mana itu merupakan kombinasi dari n-paramater, tiga dimensi {-1, 0, +1}, dengan rancangan percobaan n-parameter, dan dua dimensi {-0.5, 0.5} menghasilkan 13 titik setiap iterasi dalam dua parameter. Setelah mesin mengevaluasi semua sampel pada masing-masing iterasi kemudian memilih titik yang terbaik. (Staelin, 2003).

3.8 Confusion Matrix

Berikut ini merupakan hasil dari *confusion matrix* (Novandya, 2017)

Tabel 3.1 Tabel *Confusion Matrix*

		Prediksi	
Aktual		C1	C2
	C1	F11	F12
	C2	F21	F22

Keterangan:

F11 = jumlah prediksi yang tepat bersifat negatif (*True Negative*).

F21 = jumlah prediksi yang salah bersifat positif (*False Positive*).

F12 = jumlah prediksi yang salah bersifat negatif (*False Negative*).

F22 = Jumlah prediksi yang tepat bersifat positif (*True Positive*).

Aktual merupakan klasifikasi status hujan yang sebelumnya telah diklasifikasikan terlebih dahulu. Prediksi merupakan hasil dari klasifikasi variabel status yang dihasilkan oleh program.

Beberapa persyaratan yang telah didefinisikan untuk matrik klasifikasi di antaranya sebagai berikut:

- Accuracy* merupakan proporsi jumlah prediksi benar. Rumus akurasi adalah:

$$Akurasi = \frac{F_{11}+F_{22}}{F_{11}+F_{21}+F_{12}+F_{22}} \quad \dots (3.9)$$

- Recall* atau *true positive* (TP) adalah proporsi kasus positif yang diidentifikasi dengan benar, yang dapat dihitung dengan persamaan:

$$True\ Positive = \frac{F_{22}}{F_{12}+F_{22}} \quad \dots (3.10)$$

- Tingkat *false positive* (FP) adalah proporsi kasus negatif yang salah diklasifikasikan sebagai positif, yang dapat dihitung dengan menggunakan persamaan:

$$False\ Positive = \frac{F_{12}}{F_{11}+F_{21}} \quad \dots (3.11)$$

- d. Tingkat *true negative* (TN) didefinisikan sebagai proporsi kasus negatif yang diklasifikasikan dengan benar, dapat dihitung dengan menggunakan persamaan:

$$True\ Negative = \frac{F_{11}}{F_{11}+F_{21}} \quad \dots (3.12)$$

- e. Tingkat *false negative* (FN) adalah proporsi kasus positif yang salah diklasifikasikan sebagai negatif, yang dihitung dengan menggunakan persamaan:

$$False\ Negative = \frac{F_{21}}{F_{12}+F_{21}} \quad \dots (3.13)$$

- f. *Precision* (P) adalah proporsi prediksi kasus positif yang benar, yang dihitung dengan menggunakan persamaan:

$$P = \frac{F_{22}}{F_{21}+F_{22}} \quad \dots (3.14)$$

BAB IV

METODOLOGI PENELITIAN

4.1. Populasi dan Sampel

Populasi yang ada adalah data curah hujan harian di seluruh Indonesia tahun 2013 hingga tahun 2017, sedangkan sampel yang digunakan pada penelitian ini adalah curah hujan harian di Provinsi Jawa Timur mulai 01 Januari Tahun 2013 hingga 31 Agustus tahun 2017 pada seluruh stasiun BMKG di Jawa Timur.

4.2. Sumber Data

Data yang digunakan yaitu bersumber dari *website* resmi BMKG www.bmkg.go.id

4.3. Variabel Penelitian

Pada penelitian ini memiliki dua macam variabel yaitu variabel independen dan variabel dependen.

Tabel 4.1 Variabel Penelitian

Jenis Variabel	Nama Variabel	Skala
Variabel Independen (X)	Rata-rata Suhu (°C)	Rasio
	Rata-rata Kelembaban (%)	Nominal
	Rata-rata Lama Penyinaran (jam)	Nominal
	Rata-rata Kecepatan Angin (knot)	Rasio
Variabel Dependend (Y)	Status	Ordinal

4.3.1 Variabel Dependend

Variabel dependen yang digunakan pada penelitian ini adalah variabel status. Status merupakan hasil kategori dari variabel curah hujan yang ada di *website* BMKG yang ditentukan oleh peneliti. Peneliti memberi kategori pada variabel status berdasarkan ketentuan press releas BMKG tahun 2010. Menurut BMKG sendiri “curah hujan merupakan ketebalan air hujan yang terkumpul pada luasan 1 m². Alat pengukur hujan ada dua yaitu alat penakar biasa (manual *raingauge*) dan alat penakar otomatis (otomatis *rainguage*)”. Dimana pada variabel status terdapat beberapa kategori yaitu;

Tabel 4.2 Kategori Variabel Dependen

Variabel	Definisi Operasional Variabel		
	Hujan	Hujan Ekstrem	Tidak Hujan
Status	Jika nilai pada variabel curah hujan > 0 dan ≤ 50 mm / hari.	Jika nilai pada variabel curah hujan > 50 mm / hari	Jika nilai pada curah hujan = 0 mm / hari.

4.3.1 Variabel Independen

Pada penelitian ini variabel independen yang digunakan ada empat yaitu rata-rata suhu ($^{\circ}\text{C}$), rata-rata kelembaban (%), lama penyinaran (jam), dan rata-rata kecepatan angin (knot).

Tabel 4.3 Variabel Independen

Variabel	Definisi Operasional Variabel
Rata-rata Suhu ($^{\circ}\text{C}$)	Rata-rata suhu setiap hari dalam satuan $^{\circ}\text{C}$ yang diperoleh dari <i>website</i> bmkg.go.id
Rata-rata Kelembaban (%)	Rata-rata kelembaban setiap hari dalam persentase diperoleh dari <i>website</i> bmkg.go.id
Rata-rata Lama Penyinaran (jam)	Rata-rata lama penyinaran matahari dalam satuan jam untuk setiap hari diperoleh dari <i>website</i> bmkg.go.id
Rata-rata Kecepatan Angin (knot)	Rata-rata kecepatan angin dalam satuan knot untuk setiap hari yang diperoleh dari <i>website</i> bmkg.go.id

4.1. Metode Pengumpulan Data

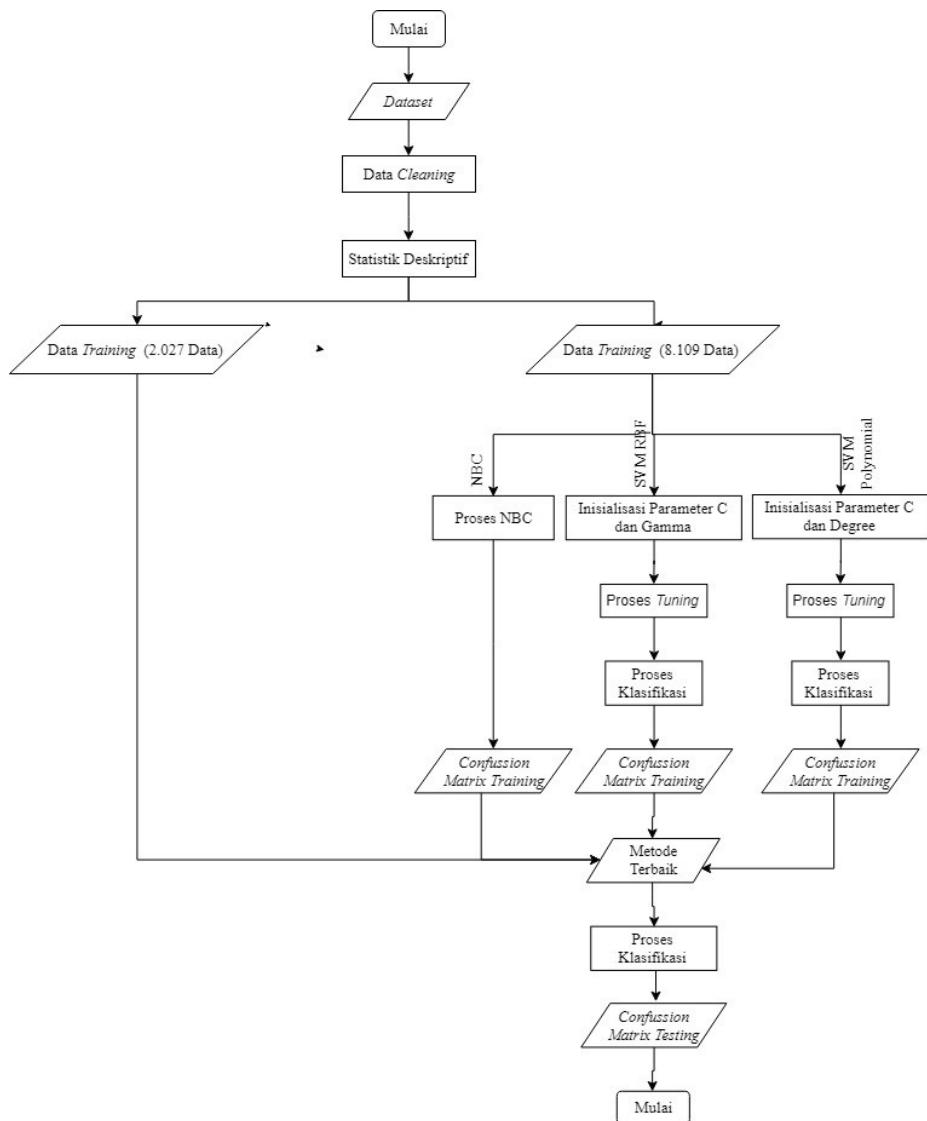
Pada penelitian ini data yang digunakan merupakan data sekunder yaitu yang diambil dari *website* resmi BMKG Indonesia, dengan beberapa data pendukung yang diambil dari *website* resmi Badan Nasional Penanggulangan Bencana (BNPB), dan Badan Pusat Statistik (BPS) Indonesia.

4.2. Metode Analisis Data

Penelitian ini menggunakan perbandingan hasil metode analisis klasifikasi *Naïve Bayes Classifier* (NBC) dan *Support Vector Machine* (SVM). Untuk metode SVM peneliti menggunakan kernel RBF dan Polynomial. SVM RBF menggunakan parameter $C = 1, 5, 10, 50$, dan 100 , sedangkan untuk $\Gamma = 1, 2, 3, 4, 5, 6$. Untuk SVM Polynomial menggunakan parameter $C = 1, 5, 10, 50$, dan 100 , sedangkan parameter Degree = 1, dan 2.

4.3. Tahapan Analisis

Adapun tahapan penelitian yang dilakukan dengan menggunakan metode NBC dan SVM adalah sebagai **gambar 4.1**.



Gambar 4.1. Flowchart Penelitian

1. Tahapan pertama adalah pengumpulan data, pengumpulan data ini kemudian terbentuk *dataset* yang akan digunakan dalam penelitian ini.
2. Data *cleaning* merupakan tahapan berikutnya setelah *dataset* terkumpul, data *cleaning* ini dilakukan untuk menghilangkan *outlier*, data missing dan variabel yang tidak digunakan.
3. Analisis deskriptif ini dilakukan pada masing-masing variabel dengan menggunakan tabel atau grafik.
4. Pada klasifikasi jenis data pada umumnya dibagi menjadi dua yaitu data *training* dan data *testing*.
5. Data *training* yaitu data latih yang digunakan, pada penelitian ini data *cleaning* yang digunakan adalah 80% dari total data yaitu sebanyak 8.109 data, sedangkan data *testing* yang digunakan sebesar 20% yaitu sebanyak 2.027. Setelah pembagian data *training* dan data *testing* dilakukan, data *training* di lakukan pada proses NBC, SVM dengan Kernel RBF, dan proses SVM dengan Kernel Polynomial.
6. Pada proses metode NBC dilakukan perhitungan *probability* baik *prior* maupun *posterior*. Proses NBC dilakukan pada masing-masing observasi. Proses tersebut dilakukan dengan menggunakan *software*. Setelah proses NBC dilakukan, maka akan didapatkan hasil tabel *confussion matrix*. Tabel tersebut digunakan untuk mencari nilai akurasi pada metode NBC.
7. Selanjutnya data *training* yang sama digunakan pada proses SVM baik kernel RBF maupun Polynomial. Pada SVM kernel RBF perlu adanya inisialisasi parameter C dan Gamma. Inisialisasi parameter ditentukan oleh peneliti berdasarkan penelitian terdahulu. Setelah nilai paramater telah ditentukan, selanjutnya adalah melakukan optimasi untuk nilai parameter dengan proses *tuning*. Langkah selanjutnya pada proses SVM kernel RBF adalah proses klasifikasi yang akan menghasilkan tabel *confussion matrix*. Tabel tersebut yang nantinya digunakan untuk menghitung akurasi.
8. Untuk proses SVM kernel Polynomial dengan menggunakan data *training* yang sama juga memerlukan inisialisasi parameter C dan Degree. Penentuan nilai parameter ditentukan berdasarkan penelitian terdahulu. Nilai parameter

telah ditentukan, maka langkah selanjutnya adalah melakukan optimasi nilai parameter. Optimasi nilai parameter biasa disebut dengan proses *tuning*. Proses *tuning* dilakukan oleh *software* yang nantinya menghasilkan nilai parameter terbaik. Setelah memperoleh parameter yang terbaik, maka selanjutnya adalah proses klasifikasi dengan SVM kernel Polynomial. Proses dilakukan oleh *software*, setelah proses klasifikasi selesai maka akan menghasilkan tabel *confussion matrix*. Tabel *confussion matrix* ini dapat digunakan untuk mencari nilai akurasi.

9. Langkah selanjutnya setelah diperoleh akurasi pada metode NBC, SVM RBF dan SVM Polynomial adalah membandingkan nilai akurasi pada data *training*. Nilai akurasi yang paling tinggi merupakan metode klasifikasi terbaik untuk data penelitian ini. Metode tersebut yang nantinya akan digunakan untuk prediksi yang akan datang.
10. Tahapan selanjutnya setelah diperoleh model dari model terbaik pada data *training*, maka akan dilakukan proses klasifikasi dengan data *testing*. Proses dilakukan dengan *software*.
11. Selanjutnya setelah proses untuk data *testing* dilakukan, maka akan menghasilkan tabel *confussion matrix* untuk data *testing*.
12. Pada data *testing* juga dilakukan proses NBC dari model yang telah diperoleh dengan data *training*, sehingga data *testing* akan menghasilkan *confussion matrix*.

BAB V

HASIL DAN PEMBAHASAN

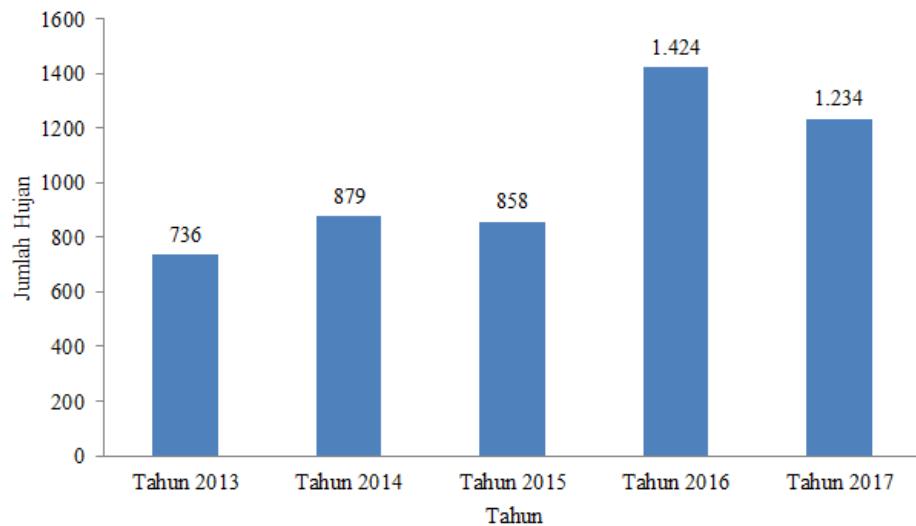
Pada bab ini akan dilakukan analisis mengenai hasil yang diperoleh berdasarkan studi kasus dan teori-teori yang telah dijelaskan pada bab-bab sebelumnya. Analisis yang digunakan adalah, analisis deskriptif, analisis perbandingan klasifikasi antara metode NBC dan SVM. Dengan menggunakan *software R.3.3.2*.

Sebelum penentuan metode klasifikasi NBC dan SVM pada penelitian ini, telah terlebih dahulu dilakukan analisis terhadap metode klasifikasi lainnya seperti, *Classification Tree*, *Artificial Neural Network*, dan *K-NN*. Metode *Classification Tree* menghasilkan nilai akurasi sebesar 77,26% untuk data *training*. Untuk metode ANN menghasilkan nilai akurasi sebesar 76,8%. Metode K-NN menghasilkan akurasi sebesar 77,1% untuk data *training*. Sehingga berdasarkan hasil akurasi dari metode yang telah dilakukan, maka dalam penelitian ini memilih dua metode dengan akurasi paling tinggi dari lima metode yang ada yaitu metode NBC dan SVM.

Variabel independen yang digunakan yaitu rata-rata suhu, rata-rata kelembaban, rata-rata lama penyinaran, dan rata-rata kecepatan angin. Data yang diperoleh sudah termasuk didalamnya berbagai karakteristik yang berpengaruh seperti arah angin, suhu tanah, dan ketinggian tempat terhadap variabel independen tersebut. Sehingga dalam proses klasifikasi dengan menggunakan variabel-variabel tersebut sudah langsung mengklasifikasi variabel tersebut tanpa adanya faktor wilayah ataupun faktor geografis.

5.1 Analisis Deskriptif

Analisis deskriptif merupakan analisis yang digunakan untuk mengetahui secara keseluruhan gambaran curah hujan di Indonesia. Pada analisis deskriptif ini menggunakan beberapa macam diagram untuk menggambarkan keadaan.



Gambar 5.1 Jumlah Status Hujan Jawa Timur Tahun 2013 – 2017

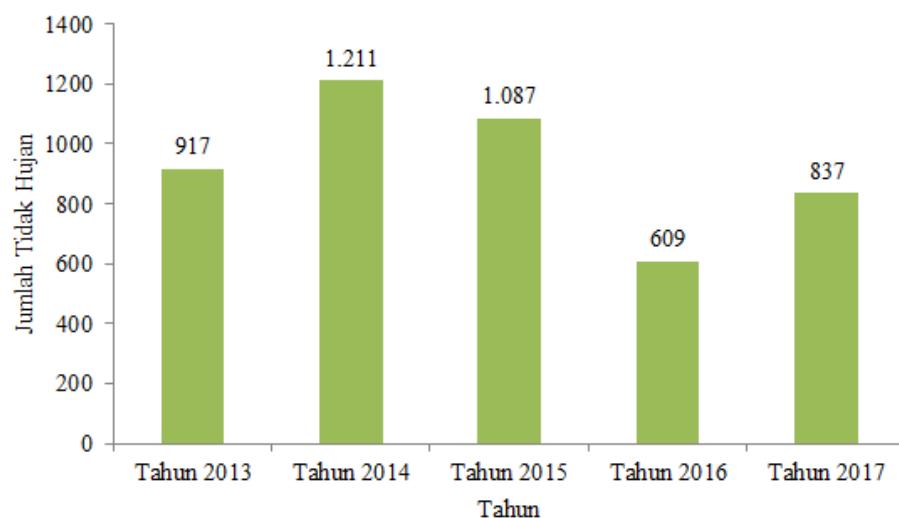
Pada **gambar 5.1** terdapat 5.136 data untuk status hujan. Tahun 2013 hujan yang terjadi di Jawa Timur ada 736 dari 5.136 data dari tahun 2013 hingga tahun 2017. Pada tahun 2014 di Jawa Timur terjadi hujan sebanyak 879, jumlah ini meningkat sebanyak 143 kejadian hujan. Untuk tahun 2015 terjadi hujan sebanyak 858, pada tahun ini mengalami penurunan sebanyak 21 kejadian dibandingkan tahun 2014. Tahun 2016 mengalami peningkatan jumlah kejadian hujan secara drastis yaitu meningkat sebanyak 566. Pada tahun 2017 terdapat penurunan sebanyak 190 yaitu dari tahun 2016 sebanyak 1.424 menjadi 1.234 pada tahun 2017.



Gambar 5.2 Jumlah Status Hujan Ekstrem Jawa Timur Tahun 2013 – 2017

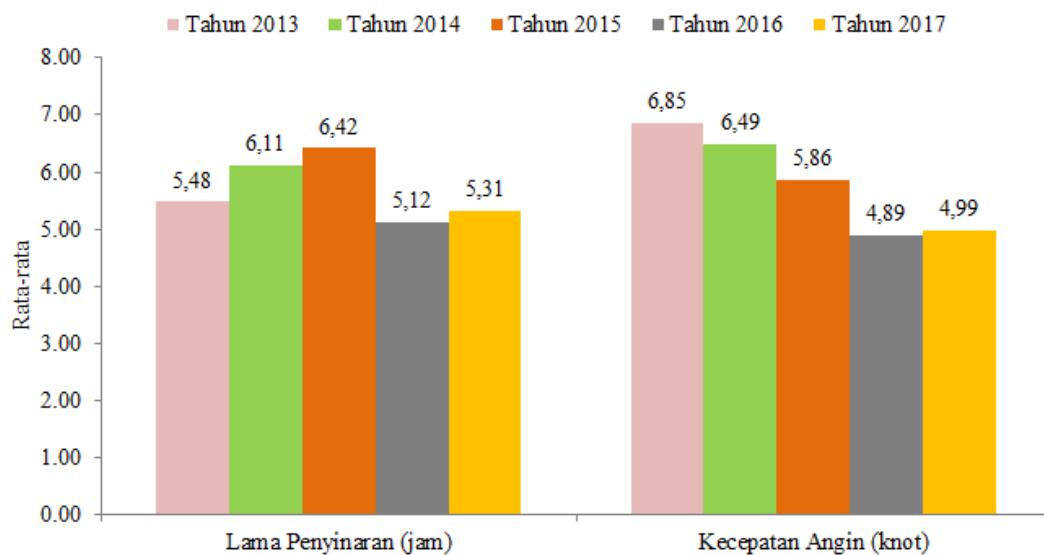
Terdapat 344 kejadian hujan ekstrem di Jawa Timur dari tahun 2013 hingga tahun 2017, seperti pada **gambar 5.2**. Pada tahun 2013 untuk kejadian hujan ekstrem terdapat 42, pada tahun ini merupakan tahun paling sedikit terjadinya jumlah kejadian hujan ekstrem diantara tahun 2014, 2015, 2016, dan 2017. Tahun 2014 mengalami peningkatan jumlah hujan ekstrem dari tahun 2013. Mulai dari tahun 2013 hingga 2015 kejadian hujan ekstrem terus mengalami kenaikan yang tidak terlalu besar yaitu 8 dan 3. Akan tetapi, pada tahun 2016 terjadi kenaikan jumlah hujan ekstrem sebesar 58 jumlah hujan ekstrem. Kenaikan jumlah kejadian hujan ekstrem ini dapat menimbulkan berbagai macam bencana seperti banjir. Terbukti pada tahun 2016 jumlah banjir meningkat dari tahun-tahun sebelumnya. Pada tahun 2016 ini juga merupakan tahun dimana jumlah hujan ekstrem banyak terjadi. Pada tahun 2017 mengalami penurunan dari tahun 2016 sebelumnya sebesar 31 jumlah kejadian hujan ekstrem.

Jumlah kejadian hujan dan hujan ekstrem paling banyak terjadi pada tahun 2016. Hal tersebut dikarenakan pada tahun 2016 Indonesia memasuki masa kemarau basah. Kemarau basah merupakan tingginya intensitas curah hujan pada saat musim kemarau. Terjadinya kemarau basah tahun 2016 disebabkan oleh tidak kuatnya Monsum Autralia (angin timuran), dan kondisi perairan di Indonesia menjadi lebih hangat (Rini, 2017).



Gambar 5.3 Jumlah Status Tidak Hujan Jawa Timur Tahun 2013 – 2017

Berdasarkan **gambar 5.3** terdapat 4.661 kejadian tidak hujan dari tahun 2013 hingga tahun 2017. Tahun 2013 terdapat 917 kejadian tidak hujan, kemudian tahun 2014 mengalami peningkatan jumlah kejadian tidak hujan sebesar 294. Untuk tahun 2015 mengalami penurunan sebesar 124 kejadian tidak hujan menurut data yang ada. Pada tahun 2016 mengalami penurunan jumlah kejadian tidak hujan sebesar 478, hal tersebut juga dapat dilihat pada **gambar 5.1** dan **gambar 5.2** jumlah kejadian hujan yang meningkat. Jika disimpulkan maka tahun 2016 ini merupakan tahun yang paling sedikit jumlah tidak hujan dari pada tahun 2013, 2014, 2015, dan tahun 2017. Pada tahun 2017 jumlah kejadian tidak hujan mengalami peningkatan sebesar 228.

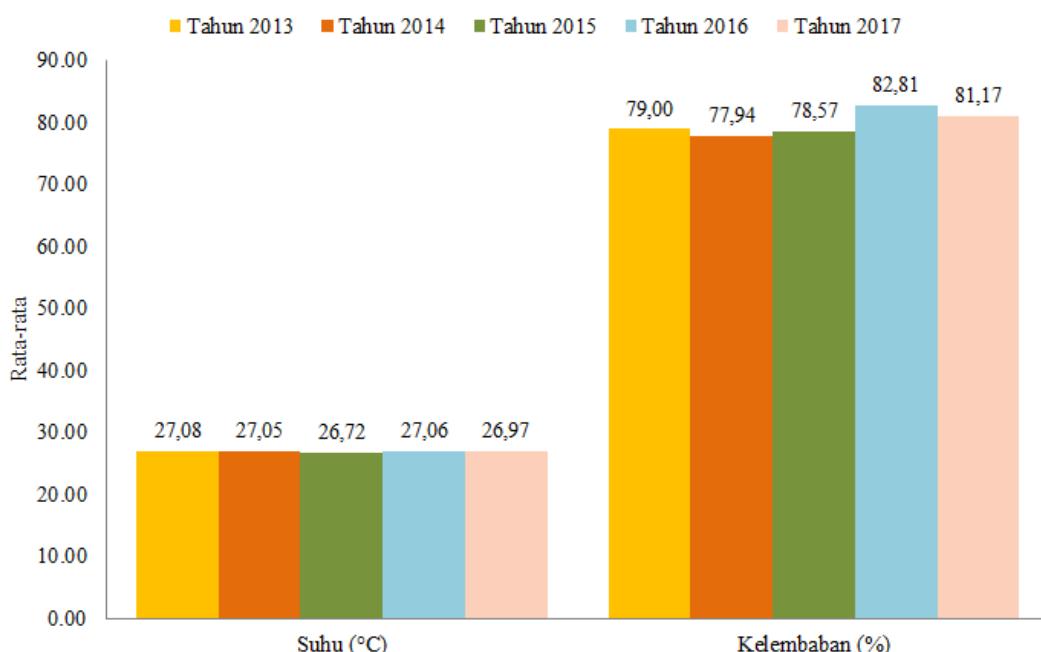


Gambar 5.4 Deskriptif Variabel Lama Penyinaran dan Kecepatan Angin

Berdasarkan **gambar 5.4** merupakan deskriptif untuk variabel rata-rata lama penyinaran dan rata-rata kecepatan. Untuk sumbu Y merupakan rata-rata masing-masing variabel sedangkan untuk sumbu X merupakan tahun yaitu dari tahun 2013 hingga tahun 2017. Untuk rata-rata lama penyinaran tertinggi pada tahun 2016 yaitu sebesar 6,4 sedangkan untuk yang paling rendah ada pada tahun 2016 yaitu sebesar 5,12. Untuk tahun 2016 merupakan tahun paling rendah rata-rata penyinaran hal tersebut terjadi karena pada tahun 2016 ini tahun dimana jumlah kejadian hujan paling banyak. Variabel lama penyinaran mengalami

perubahan yang tidak signifikan, jika dilihat dari tahun 2013 hingga tahun 2015 mengalami kenaikan, tetapi mengalami penurunan pada tahun 2016 sebesar 1,3.

Untuk variabel rata-rata kecepatan angin mengalami penurunan signifikan dari tahun 2013 hingga tahun 2016. Untuk rata-rata kecepatan angin yang paling tinggi ada pada tahun 2013 yaitu sebesar 6,85 knot. Sedangkan, untuk rata-rata kecepatan angin yang paling rendah tahun 2016 yaitu sebesar 4,89 knot. Pada tahun 2017 mengalami kenaikan rata-rata kecepatan angin yang tidak cukup besar, sedangkan dari tahun 2013 hingga tahun 2016 mengalami penurunan rata-rata kecepatan angin. Berdasarkan **gambar 5.4** penurunan rata-rata kecepatan angin paling tinggi terjadi dari tahun 2015 ke tahun 2016 sebesar 0,97 knot.



Gambar 5.5 Deskriptif Variabel Kelembaban dan Suhu

Berdasarkan **gambar 5.5** diperoleh gambaran umum untuk masing-masing variabel suhu dan kelembaban. Rata-rata suhu mengalami penurunan pada tahun 2014, 2015 dan 2016. Pada tahun 2016 rata-rata suhu mengalami kenaikan sebesar $0,34493\text{ }^{\circ}\text{C}$ dari tahun 2015. Variabel rata-rata suhu mengalami perubahan yang sangat kecil setiap tahunnya. Rata-rata suhu paling tinggi pada tahun 2013 yaitu sebesar $27,08\text{ }^{\circ}\text{C}$ sedangkan untuk rata-rata suhu paling rendah pada tahun

2015 yaitu sebesar 26,72 °C. Selisih rata-rata suhu dari suhu terbesar dan suhu terkecil adalah 0,36°C. Untuk variabel rata-rata kelembaban tahun 2014 mengalami penurunan sekitar 2,94% kemudian pada tahun 2015 mengalami kenaikan sebesar 0,63%. Rata-rata kelembaban paling tinggi pada tahun 2016 yaitu sebesar 82,8141%, dan yang paling rendah pada tahun 2014 yaitu sebesar 77,9402%. Pada umumnya variabel rata-rata kelembaban mengalami perubahan yang berfluktuatif setiap tahunnya.

5.1 Klasifikasi dengan NBC dan SVM

5.1.1 Pembagian Data *Training* dan Data *Testing*

Data pada klasifikasi umumnya dibagi menjadi dua bagian yaitu data *training* dan data *testing*. Tidak ada penentuan jumlah data *training* maupun data *testing*. Oleh karena itu, tabel 5.1 berikut merupakan pembagian data *training* dan data *testing* pada penelitian ini.

Tabel 5.1 Penggunaan Data *Training* dan Data *Testing*

Data	Jumlah
<i>Training</i>	8.109
<i>Testing</i>	2.027
Total	10.136

Berdasarkan tabel 5.1, total data yang digunakan pada penelitian ini adalah 10.136, dengan komposisi 80% untuk data *training* yaitu sebesar 8.109 data, dan sisanya 2.027 untuk data *testing*. Jumlah data *training* sebesar 80% tersebut ditentukan oleh peneliti berdasarkan penelitian sebelumnya.

5.1.2 Analisis *Naïve Bayes Classifier*

Pembagian untuk data *training* dan data *testing* telah dilakukan, maka selanjutnya adalah analisis dengan metode klasifikasi. Metode klasifikasi yang digunakan adalah NBC dimana metode NBC ini menggunakan nilai peluang dalam menentukan kelas klasifikasi. Analisis menggunakan NBC dengan jumlah data 8.109 nantinya akan menghasilkan kelas klasifikasi dan tabel prediksi. Kelas klasifikasi yang dihasilkan adalah hujan, hujan ekstrem, dan tidak hujan. Berikut merupakan hasil pembahasan dengan menggunakan metode NBC.

Tabel 5.2 *Priori Probability*

Hujan	Hujan Ekstrem	Tidak Hujan
0,51079	0,03404	0,45517

Berdasarkan tabel 5.2 diperoleh nilai probabilitas awal untuk masing-masing status pada variabel status hujan, hujan ekstrem, dan tidak hujan. Total untuk prediksi masing-masing status hujan dapat dilihat pada tabel 5.3. Untuk probabilitas hujan yaitu sebesar 0.51079 diperoleh dari 4.142 dibagi dengan 8.109. Probabilitas prediksi hujan ekstrem sebesar 0.03404 diperoleh dari 276 dibagi dengan 8.109. Untuk probabilitas prediksi status tidak hujan sebesar 0.45517 diperoleh dari 3.691 dibagi dengan 8.109. *Priori probability* merupakan tahapan untuk mencari nilai probabilitas pada masing-masing pengamatan yang nantinya akan menghasilkan klasifikasi.

Tabel 5.3 Hasil Confussion Matrix NBC Data Training

Observasi	Prediksi			Total
	Hujan	Hujan Ekstrem	Tidak Hujan	
Hujan	3.380	253	726	4.358
Hujan Ekstrem	12	8	3	24
Tidak Hujan	750	15	2.962	3.730
Total	4.142	276	3.691	8.109

Tabel 5.3 merupakan hasil perbandingan antara prediksi dengan observasi untuk data *training*. Pada klasifikasi dikatakan benar jika pada variabel status yaitu hujan, hujan ekstrem dan tidak hujan dalam observasi sama dengan dalam prediksi. Pada variabel status hujan diperoleh tepat klasifikasi sebesar 3.380 kejadian. Pada observasi yang termasuk pada status hujan ekstrem tetapi pada prediksi masuk kedalam hujan ada 12, sedangkan yang pada observasi masuk dalam klasifikasi tidak hujan tetapi pada prediksi masuk pada klasifikasi hujan ada 750. Untuk klasifikasi hujan ekstrem yang tepat klasifikasi terdapat 8 kejadian. Terdapat 253 untuk yang salah klasifikasi, dimana dalam data observasi termasuk pada klasifikasi hujan, tetapi dalam hasil prediksi masuk dalam klasifikasi hujan ekstrem, dan dalam prediksi masuk pada klasifikasi hujan ekstrem tetapi pada observasi masuk pada klasifikasi tidak hujan ada 726. Untuk status tidak hujan yang tepat klasifikasi ada 2.962.

Total untuk data observasi pada status hujan ada 4.358, pada status hujan ekstrem ada 24, dan pada status tidak hujan ada 3.730. Total untuk data prediksi pada status hujan ada 4.142, pada status hujan ekstrem ada 276, sedangkan pada status tidak hujan ada 3.691. Jumlah seluruh data yang digunakan ada 8.109 data. Berdasarkan tabel 5.3 maka dapat diperoleh tingkat akurasi dan *error*.

$$\text{Akurasi} = \frac{\sum(\text{prediksi benar})}{\sum(\text{semua prediksi})} \dots (5.1)$$

$$= \frac{3.380+8+2.962}{8.109} = 0,78308$$

$$\text{Error} = \frac{\sum(\text{prediksi salah})}{\sum(\text{semua prediksi})} \dots (5.2)$$

$$= \frac{253+726+12+3+750+15}{8.109} = 0,21692$$

Berdasarkan hasil akurasi pada data *training* yaitu persamaan 5.1 diperoleh nilai akurasi sebesar 0,78308 atau sebesar 78,308%. hasil akurasi tersebut merupakan hasil akurasi yang cukup besar, akan tetapi karena pada penelitian ini peneliti ingin membandingkan nilai akurasi pada metode NBC dan SVM, maka untuk mendapatkan prediksi dari model dipilih nilai akurasi yang besar dari kedua metode tersebut. Untuk data *testing* dengan metode NBC memperoleh akurasi sebesar 0,77898 atau 77,898%.

5.1.3 Analisis Klasifikasi *Support Vector Machine*

Metode SVM bekerja dengan cara mencari *hyperplane* atau garis pemisah antar klasifikasi. Karena data yang digunakan merupakan data kontinu dan data tidak linear maka metode SVM yang tepat adalah dengan Kernel Gaussian. Algoritma SVM dengan Kernel Gaussian merupakan pendekatan untuk data kontinu yang tidak linear. Berikut merupakan hasil dari klasifikasi metode SVM dengan metode kernel RBF pada data *training*.

Tabel 5.4 Akurasi SVM dengan Kernel RBF Data *Training*

C	Gamma					
	1	2	3	4	5	6
1	0,80602	0,81379	0,81958	0,82723	0,83129	0,83820
5	0,81181	0,82599	0,83894	0,85843	0,87372	0,88469
10	0,81292	0,83339	0,85461	0,87569	0,89185	0,90504
50	0,82304	0,85683	0,88987	0,91257	0,93229	0,94759
100	0,82809	0,86977	0,90196	0,92823	0,94858	0,96251

Berdasarkan tabel 5.4 diperoleh untuk nilai parameter C dan Gamma pada metode SVM. Jika dilihat pada tabel 5.4 semakin tinggi nilai C dan semakin tinggi nilai Gamma maka nilai akurasi yang diperoleh akan semakin tinggi. Untuk nilai C merupakan nilai positif yaitu ≥ 0 , sedangkan untuk nilai Gamma adalah ≥ 0 . Untuk nilai C=100 dan Gamma=6 merupakan parameter yang menghasilkan nilai akurasi paling besar, sedangkan untuk C=1 dan Gamma=1 merupakan parameter yang menghasilkan nilai akurasi paling kecil. Berikut ini merupakan tabel hasil akurasi SVM dengan kernel RBF untuk data *testing*.

Tabel 5.5 Akurasi SVM dengan Kernel RBF Data *Testing*

C	Gamma					
	1	2	3	4	5	6
1	0.80365	0.80760	0.80365	0.80414	0.80217	0.79674
5	0.80661	0.80266	0.79329	0.78688	0.77553	0.76221
10	0.80316	0.80168	0.78737	0.77356	0.75678	0.75481
50	0.79921	0.78096	0.74840	0.73458	0.71880	0.70597
100	0.79970	0.76184	0.73458	0.71436	0.70992	0.69610

Berdasarkan tabel 5.5 nilai akurasi dengan menggunakan data *testing* pada metode SVM kernel RBF diperoleh nilai akurasi paling kecil adalah C=100 dan Gamma=6. Hal tersebut berlawanan dengan hasil akurasi pada data *training* tabel 5.4. Oleh karena itu, diperlukan metode optimasi untuk mencari nilai parameter terbaik. Untuk mencari nilai optimasi dari parameter C dan Gamma pada penelitian ini menggunakan proses *tuning*. Pada proses *tuning* tersebut diperoleh nilai optimasi untuk parameter C adalah 1 dan untuk Gamma adalah 1. Sehingga, pada penelitian ini dengan menggunakan Kernel RBF diperoleh parameter yang

terbaik adalah C=1 dan Gamma=1. Untuk nilai akurasi pada nilai C= 1 dan Gamma= 1 yaitu sebesar 0,80602.

Tabel 5.6 Confussion Matrix SVM dengan Kernel RBF (C= 1, Gamma= 1)

Observasi	Prediksi			Total
	Hujan	Hujan Ekstrem	Tidak Hujan	
Hujan	3.628	267	783	4.678
Hujan Ekstrem	0	0	0	0
Tidak Hujan	514	9	2.908	3.431
Total	4.142	276	3.691	8.109

Tabel 5.6 merupakan hasil klasifikasi menggunakan metode SVM dengan C= 1 dan Gamma= 1 untuk data *training*. Pada metode ini untuk status hujan pada observasi mendapatkan tepat klasifikasi yaitu sebesar 3.628 data, dan tepat klasifikasi pada status hujan ekstrem sebesar 0 data, sedangkan pada status tidak hujan tepat klasifikasi sebesar 2.691 data. Tidak ada data klasifikasi yang salah terjadi pada observasi hujan ekstrem sedangkan pada prediksi masuk klasifikasi hujan. Klasifikasi yang salah lainnya terjadi pada observasi hujan tetapi pada prediksi masuk pada hujan ekstrem yaitu ada 267 data. Terdapat 514 data yang salah klasifikasi yang seharusnya masuk pada tidak hujan tetapi pada prediksi masuk klasifikasi hujan. Kesalahan pada klasifikasi inilah yang membuat *error* itu ada.

$$\text{Akurasi} = \frac{\Sigma(\text{prediksi benar})}{\Sigma(\text{semua prediksi})} \quad \dots(5.3)$$

$$= \frac{3.628+0+2.908}{8.109} = 0,80602$$

$$\text{Error} = \frac{\Sigma(\text{prediksi salah})}{\Sigma(\text{semua prediksi})} \quad \dots(5.4)$$

$$= \frac{267+783+514+9}{8.109} = 0,19398$$

Pada persamaan 5.3 diperoleh hasil akurasi untuk metode SVM sebesar 0,806018 atau 80,6018%. Semakin besar nilai akurasi suatu metode maka, semakin bagus metode tersebut untuk digunakan pada data. Bagus dalam arti antara observasi dengan prediksi sudah tepat klasifikasi. Untuk nilai *error* yang diperoleh sebesar 0,19398 atau 19,398%.

Tabel 5.7 Akurasi SVM dengan Kernel Polynomial Data *Training*

C	Degree	
	d1	d2
1	0.78986	0.58244
5	0.79011	0.58281
10	0.79011	0.58281
50	0.79011	0.58244
100	0.78999	0.58244

Berdasarkan tabel 5.7 hasil akurasi metode SVM dengan kernel Polynomial untuk data *training* yang paling besar adalah dengan nilai parameter C=5, C=10, dan C=50 dengan Degree=1. Untuk akurasi yang paling kecil semua nilai parameter C untuk Degree=1. Dengan menggunakan *tuning* diperoleh parameter yang paling optimal adalah C=5 dan Degree=1.

Tabel 5.8 Akurasi SVM dengan Kernel Polynomial Data *Testing*

C	Degree	
	d1	d2
1	0,79477	0,56043
5	0,79428	0,55994
10	0,79526	0,55994
50	0,79477	0,55994
100	0,79477	0,55994

Tabel 5.8 merupakan hasil akurasi untuk data *testing*, dengan nilai akurasi tebesar pada C=1, 50, dan 100 sedangkan Degree=1. Berdasarkan tabel 5.7 diperoleh nilai parameter yang optimal dengan menggunakan *tuning* C=5 dan Degree=1, oleh karena itu pada data *testing* akan dipilih nilai parameter C=5 dan Degree=1. Model yang terbentuk berdasarkan data *training*, oleh karena itu proses tuning hanya dilakukan pada data *training*. Berdasarkan pada tabel 5.6 diperoleh untuk nilai akurasi tertinggi pada SVM dengan kernel Polynomial yaitu dengan

nilai C paling kecil 5 dan Degree = 1 yaitu sebesar 0,79011. Untuk data *testing* diperoleh nilai akurasi pada C=5 dan Degree=1 sebesar 0,79428. Untuk nilai akurasi SVM kernel Polynomial lebih kecil dari pada SVM kernel RBF.

Tabel 5.9 *Confusion Matrix* SVM dengan Kernel Polynomial (C= 5, Degree= 1)

Observasi	Prediksi			Total
	Hujan	Hujan Ekstrem	Tidak Hujan	
Hujan	3.525	267	809	4.601
Hujan Ekstrem	0	0	0	0
Tidak Hujan	617	9	2.882	3.499
Total	4.142	276	3.691	8.109

Tabel 5.9 merupakan tabel SVM kernel Polynomial dengan nilai C = 5 dan Degree = 1 untuk data *training*. Berdasarkan tabel 5.7 maka dapat diperoleh nilai akurasi dan *error*. Berikut merupakan hasil nilai akurasi dan *error*.

$$\text{Akurasi} = \frac{\Sigma(\text{prediksi benar})}{\Sigma(\text{semua prediksi})} \quad \dots(5.5)$$

$$= \frac{3525+0+2882}{8109} = 0,79011$$

$$\text{Error} = \frac{\Sigma(\text{prediksi salah})}{\Sigma(\text{semua prediksi})} \quad \dots(5.6)$$

$$= \frac{267+809+617+9}{8109} = 0,20989$$

5.2 Perbandingan Hasil Metode NBC dan SVM

Perbandingan hasil metode NBC dan SVM ini digunakan untuk menentukan metode yang terbaik. Dalam menentukan metode yang terbaik dilihat berdasarkan nilai akurasi yang tinggi. Hasil dari kedua metode diperoleh nilai akurasi kedua metode sebagai berikut.

Tabel 5.10 Hasil Perbandingan

	SVM RBF (C=1, Gamma=1)	SVM <i>Polynomial</i> (C=5, Gamma=1)	NBC
Akurasi	0,80602	0,79011	0,78308

Pada tabel 5.10 dapat dilihat bahwa untuk hasil perbandingan nilai akurasi antara metode NBC dan SVM yang paling tinggi adalah 0,80602 yaitu metode SVM dengan menggunakan fungsi kernel RBF ($C=1$, $\Gamma=1$). Oleh karena itu, dapat disimpulkan bahwa metode SVM dengan kernel RBF ($C=1$, $\Gamma=1$) yang paling baik digunakan pada data tersebut. Metode dengan kernel RBF ($C=1$, $\Gamma=1$) yang kemudian untuk dijadikan prediksi curah hujan di masa yang akan datang. Jadi, berikut merupakan hasil klasifikasi untuk data *testing* pada metode SVM.

Tabel 5.11 *Confusion Matrix Data Testing* SVM RBF ($C=1$, $\Gamma=1$)

Observasi	Prediksi			Total
	Hujan	Hujan Ekstrem	Tidak Hujan	
Hujan	868	66	209	1.143
Hujan Ekstrem	0	0	0	0
Tidak Hujan	121	2	761	884
Total	989	68	970	2.027

Tabel 5.9 data *testing* yang digunakan berjumlah 2.027 data. Jika dilihat terdapat 868 data yang tepat klasifikasi pada status Hujan, sedangkan tepat klasifikasi pada status Hujan Ekstrem ada 0, dan untuk status Tidak Hujan tepat klasifikasi ada 761 data. Terdapat 398 data yang salah klasifikasi. Berdasarkan hasil tersebut diperoleh nilai akurasi dan *error* sebagai berikut:

$$\text{Akurasi} = \frac{\sum(\text{prediksi benar})}{\sum(\text{semua prediksi})} \quad \dots(5.7)$$

$$= \frac{868+0+761}{2.027} = 0,80365$$

$$\text{Error} = \frac{\sum(\text{prediksi salah})}{\sum(\text{semua prediksi})} \quad \dots(5.8)$$

$$= \frac{66+209+121+2}{2.027} = 0,19635$$

Berdasarkan pada persamaan 5.7 nilai akurasi untuk data *testing* dengan metode SVM kernel RBF ($C=1$, $\Gamma=1$) sebesar 80,365%, dan error sebesar

19.635%. Untuk nilai akurasi pada data *training* dengan data *testing* tidak jauh berbeda, artinya model yang diperoleh dapat digunakan untuk data baru yang serupa.

BAB VI

PENUTUP

6.1. Kesimpulan

Berdasarkan pembahasan maka diperoleh beberapa kesimpulan bahwa dari hasil analisis yang telah dilakukan, terdapat beberapa informasi yaitu:

1. Pada umumnya kondisi curah hujan di Jawa Timur pada tahun 2013 - 2017 mengalami perubahan yang berfluktuatif.
 - a. Status hujan yaitu dengan curah hujan > 0 dan ≤ 50 mm / hari paling banyak terjadi ada pada tahun 2016, sedangkan yang paling sedikit terjadi pada tahun 2013. Jumlah kejadian hujan mulai tahun 2013 – 2017 ada sebanyak 5.131 kali hujan.
 - b. Untuk status hujan ekstrem atau curah hujan harian > 50 mm/ hari paling banyak terjadi pada tahun 2016, dan untuk kejadian hujan ekstrem paling sedikit tahun 2013. Pada tahun 2013 kejadian hujan ekstrem mengalami peningkatan yang cukup tinggi yaitu sebesar 55 kejadian hujan ekstrem.
 - c. Status tidak hujan yaitu dengan curah hujan 0 mm / hari dari tahun 2013 – 2017 terdapat 4.661 kejadian tidak hujan. Kejadian tidak hujan paling banyak pada tahun 2014, sedangkan paling sedikit kejadian tidak hujan pada tahun 2016. Tahun 2016 merupakan tahun paling banyak kejadian hujan, oleh karena itu kejadian tidak hujannya paling sedikit.
2. Berdasarkan analisis pembahasan diperoleh hasil akurasi untuk data *training* dan data *testing* sebagai berikut:
 - a. Untuk metode NBC dengan data *training* sebesar 8.109 diperoleh akurasi sebesar 78,308%, sedangkan dengan data *testing* 77,898%.
 - b. Metode SVM menggunakan fungsi kernel RBF dengan C=1 dan Gamma=1 diperoleh akurasi yang tepat untuk data *training* sebesar

80,602%. Akurasi dengan data *testing* C=1 dan Gamma=1 sebesar 80,365%.

- c. Metode SVM menggunakan fungsi kernel Polynomial dengan C=5 dan Degree=1 diperoleh akurasi pada data *training* sebesar 79,011%. Sedangkan, untuk akurasi dengan data *testing* C=5 dan Degree=1 sebesar 79,428%.
3. Berdasarkan hasil akurasi pada data *training* yang diperoleh, maka hasil perbandingan nilai akurasi tersebut pada metode NBC dan SVM yang tertinggi adalah metode SVM menggunakan kernel RBF dengan C=1 dan Gamma=1. Hasil untuk akurasi data *testing* adalah sebesar 80,365%. Metode SVM inilah yang nantinya akan digunakan untuk prediksi data selanjutnya.

6.1. Saran

Saran yang dapat diberikan dari hasil penelitian ini adalah :

1. Pada penelitian selanjutnya diharapkan dapat menggunakan metode klasifikasi lain atau menggabungkan beberapa metode klasifikasi lain, sehingga diharapkan dapat meningkatkan akurasi.
2. Untuk BMKG dapat mempertimbangkan hasil prediksi klasifikasi yang ada pada penelitian ini, sehingga dapat memberikan *warning alarm* hujan ekstrem yang dapat mengakibatkan banjir.
3. Untuk Pemerintah Jawa Timur hendaknya memperhatikan prediksi kejadian hujan ekstrem, sehingga dapat melakukan upaya yang dapat mengantisiasi kejadian banjir akibat hujan eksrem, salah satunya seperti memberi pos-pos pada daerah rawan banjir untuk berjaga ketika terjadi hujan ekstrem, sehingga dalam proses evakuasi ketika hujan ekstrem dapat berjalan dengan cepat.

DAFTAR PUSTAKA

- Assaffat, Luqman. 2015. Analisis Akurasi Support Vector Machine Dengan Fungsi Kernel Gaussian RBF Untuk Prakiraan Beban Listrik Harian Sektor Industri. *Jurnal Momentum*. Vol. 11 (2)
- Artika, Yudha. 2008. *Studi Intensitas Curah Hujan Maksimum Terhadap Kemampuan Drainase Perkotaan (Studi Kasus Jl Ir H Juanda Kota Bandung)*. Skripsi. Fakultas Ilmu dan Teknologi Kebumian. Institut Teknologi Bandung. Bandung
- BMKG. 2017. *Bidang Meteorologi dan Klimatologi Geofisika*. Tersedia di <http://www.bmkg.go.id/> di akses pada 10 November 2018.
- BNPB. 2017. *Badan Nasional Penanggulangan Bencana*. Tersedia di <https://www.bnrb.go.id/home> diakses pada tanggal 10 November 2018.
- BPS. 2017. Bada Pusat Statistik. Tersedia di <https://www.bps.go.id/> diakses pada tanggal 10 November 2018.
- Estoatnowo, Ditya. 2016. *Klasifikasi Status Kelulusan Mahasiswa Menggunakan Metode CHAID dan Algoritma C4.5 (Studi Kasus : Mahasiswa Jurusan Statistika Matematika dan Ilmu Pengetahuan Alam Universitas Islam Indonesia)*. Skripsi. Fakultas Matematika dan Ilmu Pengetahuan Alam. Universitas Islam Indonesia. Yogyakarta
- F. Arfiana. 2014. *Klasifikasi Kendaraan Roda Empat Menggunakan Metode Naïve Bayes*. Skripsi. Universitas Widyaatama. Bandung
- Fibrianda, Mercury Fluorida, dan Adhitya Bhawiyuga. 2017. Analisis Perbandingan Akurasi Deteksi Serangan Pada Jaringan Komputer Dengan Metode Naïve Bayes dan Suport Vector Machine (SVM). *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, Vol. 2 (9).
- Harafani, Hani dan Romi Satria Wahono. 2015. Optimasi Parameter pada Support Vector Machine Berbasis Algoritma Genetika untuk Estimasi Kebakaran Hutan. *Journal of Intelligent System*. Vol 1 (2)
- Hidayatullah, Ade Ihsan. 2017. *Algoritma Backpropagation Untuk Prediksi Delay Pesawat Akibat Cuaca (Studi Kasus : Bandara Internasional Soekarno-*

Hatta). Skripsi. Fakultas Matematika dan Ilmu Pengetahuan Alam. Universitas Islam Indonesia. Yogyakarta

Mujiasih, Subekti. 2011. Pemanfaatan Data Mining Untuk Prakiraan Cuaca. *Jurnal Meteorologi dan Geofisika* Vol 12 (2), hlm 189:195;

Novandy, Andhika dan Oktria, Isni. 2017. Penerapan Algoritma Klasifikasi Data Mining C4.5 Pada Dataset Cuaca Wilayah Bekasi. *Jurnal* Vol 6 (2) hlm 2089 – 5615

Nugroho, Anto Satrio, Arief Budi Witarto, dan Dwi Handoko. 2003. *Support Vector Machine Teori dan Aplikasinya dalam Bioinformatika*.

Octaviani, Puspita Anna, Yuciana Wilandari, dan Dwi Ispriyanti. 2014. Penerapan Metode Klasifikasi *Support Vector Machine* (SVM) Pada Data Akreditasi Sekolah Dasar (SD) Di Kabupaten Magelang. *Jurnal Gaussian*, Vol 3 (4) ISSN: 2339-2541

Olaiya, Folorunsho. 2012. Application of Data Mining Technique in Weather Prediction and Climate Change Studies. *Journal I.J. Information Engineering and Electronic Business*, Vol 1 hlm 51-59

Pratama, Arif, Randy Cahya Wihandika, dan Dian Eka Ratnawati. 2017. Implementasi Algoritma *Support Vector Machine* (SVM) untuk Prediksi Ketepatan Waktu Kelulusan Mahasiswa. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, Vol. 2 (4).

Pratiwi, Yulinda Rizky. 2017. *Analisis Sentimen Pada Pertaite Melalui Jejaring Sosial Twitter dengan Menggunakan Metode Support Vector Machine dan Maximum Entropy*. Skripsi. Fakultas Matematika dan Ilmu Pengetahuan Alam. Universitas Islam Indonesia. Yogyakarta

Prawaka, Fanny. 2016. *Analisis Data Curah Hujan Yang Hilang Dengan Menggunakan Metode Normal Ratio, Inversed Square Distaance, dan Rata-rata Aljabar*. Skripsi. Fakultas Teknik. Universitas Lampung : Lampung.

Radhika, Y dan M Sishi. 2009. Atmospheric Temperature Prediction Using Support Vector Machines. International Journal of Computer Theory and Engineering. Vol 1 (1). Pp 1789-8201

Rani, Larissa Navia. 2015. Klasifikasi Nasabah Menggunakan Algoritma C4.5 Sebagai Dasar Pemberian Kredit. *Jurnal KomTekInfo Fakultas Ilmu Komputer* Vol:2 (2) .

- Rini, Dwi. 2017. Kilas Balik Kejadian Cuaca, Iklim, dan Gempa Bumi 2016. Tersedia di <http://www.bmkg.go.id/berita/?p=kilas-balik-kejadian-cuaca-iklim-dan-gempabumi-indonesia-rentan-bencana&lang=ID>. Diakses pada tanggal 01 Maret 2018.
- Santosa, Budi. Data Mining : *Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Graha Ilmu. Yogyakarta
- Subhan, Ahmad. dan Ahmad Zainul Fanani. 2017. *Penerapan Data Mining Untuk Menentukan Potensi Hujan Harian Dengan Menggunakan Algoritma Naïve Bayes*. Skripsi. Fakultas Teknik Informatika. Universitas Dian Nuswantoro. Semarang
- Siregar, Aplia Belina. 2017. *Pemodelan support Vector Machine Untuk Klasifikasi Curah Hujan Bulanan Di Kabupaten Indramayu*. Skripsi. Statistika. Institut Pertanian Bogor.
- Staelin, Carl. 2003. Parameter Selection for Support Vector Machine. Approved For External Publication. Israel
- Sucipto, Adi dan Akhmad Khanif Zyen. 2017. Pengembangan Model *Support Vector Machine (SVM)* Dengan Memperbanyak Dataset Untuk Prediksi Bisnis *Forex* Menggunakan Metode Kernel *Trick*. *Jurnal Informatika Upgris*. Vol.3 (1)
- Susilowati, dan Ilyas Sadad. 2015. Analisa Karakteristik Curah Hujan Di Kota Bandar Lampung. *Jurnal Kontruksia*. Vol 7 (1)

LAMPIRAN

Lampiran 1. Data Penelitian

No	Rata-rata Suhu (°C)	Rata-rata Kelembaban (%)	Rata-rata Penyinaran (jam)	Rata-rata Kecepatan Angin (knot)	Status
1	26,7	86	1,5	1	Hujan
2	24,4	87	1,2	4	Hujan
3	27,9	84	5,1	8	Hujan Ekstrem
4	27,6	85	2	11	Hujan
5	26,7	96	1,3	14	Hujan Ekstrem
6	26,7	87	4	2	Hujan
7	23,5	88	2,5	3	Hujan
8	28,1	86	6,6	8	Hujan
9	27,1	95	0,5	10	Hujan
10	27,7	86	5	12	Hujan Ekstrem
...
...
...
...
...
...
...
10127	27,2	86	2	5	Hujan
10128	28,4	80	3,8	5	Tidak Hujan
10129	27,6	87	2,7	7	Tidak Hujan
10130	26,2	86	3	1	Hujan
10131	25,4	93	5,1	3	Hujan
10132	22,2	92	0	4	Hujan
10133	28,6	81	2,5	4	Hujan
10134	28,6	83	4,2	5	Hujan
10135	23,9	85	2,2	6	Hujan
10136	28,4	83	5,3	9	Hujan

Lampiran 2. Syntax NBC

```

#-----Naive NEW-----
data=read.delim("clipboard") #all #menentukan sampel
samplesize=round(0.8*nrow(data))
set.seed(12345)
index = sample( seq_len ( nrow ( data ) ), size = samplesize )
training = data[ index, ]
View(training)
A=nrow(training)
A
testing=data [-index,]
#----- Model -----
library(naivebayes)
naive=naive_bayes(status~,data = training,usekernel = T)
naive
plot(naive)
nb=prop.table(training$Rata.rata.Kecepatan.Angin..knot.,training$data)
nb
#----- Data Training -----
xx <- training
pred22 <- predict(naive,xx)
cm22 <- table(pred22,xx$status)
accuracy22 <- sum(cm22[1,1]+cm22[2,2]+cm22[3,3])/sum(cm22)
accuracy22
list(Matrix_Uji=cm22,Akurasi=accuracy22,Total_data=sum(cm22))

#----- Data Testing -----
xtest <- testing

```

```
predtest <- predict(naive,xtest)
cmtest <- table(predtest,xtest$status)
accuracytest <- sum(cmtest[1,1]+cmtest[2,2]+cmtest[3,3])/sum(cmtest)
accuracytest
list(Matrix_Uji=cmtest,Akurasi=accuracytest,Total_data=sum(cmtest))
```

Lampiran 3. Output NBC

a. Hasil Model

```
===== Naive Bayes =====
Call:
naive_bayes.formula(formula = status ~ ., data = training, usekernel = T)

A priori probabilities:

          Hujan  Hujan Ekstrem  Tidak Hujan
0.51079048    0.03403626    0.45517326

Rata.rata.Suhu...C.      Hujan  Hujan Ekstrem  Tidak Hujan
18,5 0.0000000000 0.0000000000 0.0000000000
19  0.0000000000 0.0000000000 0.0002709293
19,4 0.0000000000 0.0000000000 0.0000000000
19,5 0.0000000000 0.0000000000 0.0002709293
19,6 0.0000000000 0.0000000000 0.0005418586
19,7 0.0000000000 0.0000000000 0.0005418586
19,8 0.0000000000 0.0000000000 0.0005418586
19,9 0.0002414293 0.0000000000 0.0002709293
20  0.0000000000 0.0000000000 0.0002709293
20,1 0.0004828585 0.0000000000 0.0005418586
20,2 0.0002414293 0.0036231884 0.0005418586
20,3 0.0007242878 0.0000000000 0.0005418586
20,4 0.0009657170 0.0036231884 0.0002709293
20,5 0.0007242878 0.0036231884 0.0013546464
20,6 0.0002414293 0.0036231884 0.0029802222
20,7 0.0012071463 0.0036231884 0.0005418586
20,8 0.0009657170 0.0036231884 0.0010837171
20,9 0.0009657170 0.0000000000 0.0008127879
21  0.0016900048 0.0000000000 0.0016255757
21,1 0.0028971511 0.0072463768 0.0024383636
21,2 0.0016900048 0.0072463768 0.0013546464
21,3 0.0041042974 0.0036231884 0.0029802222
21,4 0.0021728634 0.0036231884 0.0005418586
21,5 0.0026557219 0.0072463768 0.0021674343
21,6 0.0033800097 0.0072463768 0.0035220807
21,7 0.0031385804 0.0036231884 0.0024383636
21,8 0.0026557219 0.0036231884 0.0029802222
21,9 0.0021728634 0.0000000000 0.0035220807
22  0.0021728634 0.0072463768 0.0046057979
22,1 0.0033800097 0.0000000000 0.0048767272
22,2 0.0021728634 0.0036231884 0.0029802222
22,3 0.0045871560 0.0108695652 0.0046057979
22,4 0.0055528730 0.0108695652 0.0029802222
22,5 0.0041042974 0.0108695652 0.0051476565
22,6 0.0041042974 0.0144927536 0.0035220807
22,7 0.0033800097 0.0072463768 0.0054185857
22,8 0.0060357315 0.0000000000 0.0032511514
22,9 0.0065185901 0.0108695652 0.0059604443
```

Rata.rata.Penyinaran..jam.	Hujan	Hujan Ekstrem	Tidak Hujan
0	0.0736359247	0.1086956522	0.0083988079
0,1	0.0057943023	0.0108695652	0.0002709293
0,2	0.0123128923	0.0217391304	0.0016255757
0,3	0.0111057460	0.0108695652	0.0018965050
0,4	0.0074843071	0.0108695652	0.0008127879
0,5	0.0188314824	0.0289855072	0.0016255757
0,6	0.0101400290	0.0144927536	0.0018965050
0,7	0.0050700145	0.0108695652	0.0002709293
0,8	0.0082085949	0.0036231884	0.0010837171
0,9	0.0045871560	0.0072463768	0.0005418586
1	0.0219700628	0.0289855072	0.0021674343
1,1	0.0048285852	0.0036231884	0.0002709293
1,2	0.0137614679	0.0144927536	0.0010837171
1,3	0.0067600193	0.0108695652	0.0018965050
1,4	0.0045871560	0.0036231884	0.0010837171
1,5	0.0212457750	0.0108695652	0.0032511514
1,6	0.0043457267	0.0072463768	0.0002709293
1,7	0.0055528730	0.0036231884	0.0010837171
1,8	0.0077257364	0.0108695652	0.0008127879
1,9	0.0041042974	0.0036231884	0.0005418586
10	0.0019314341	0.0000000000	0.0243836359
10,1	0.0012071463	0.0000000000	0.0062313736
10,2	0.0009657170	0.0000000000	0.0121918179
10,3	0.0019314341	0.0000000000	0.0073150908
10,4	0.0002414293	0.0000000000	0.0054185857
10,5	0.0004828585	0.0000000000	0.0111081008
10,6	0.0002414293	0.0000000000	0.0035220807
10,7	0.0002414293	0.0000000000	0.0035220807
10,8	0.0004828585	0.0000000000	0.0029802222
10,9	0.0000000000	0.0000000000	0.0016255757
11	0.0000000000	0.0000000000	0.0021674343
11,1	0.0000000000	0.0000000000	0.0002709293
11,2	0.0004828585	0.0000000000	0.0002709293
11,3	0.0000000000	0.0000000000	0.0000000000
11,5	0.0000000000	0.0000000000	0.0005418586
2	0.0231772091	0.0253623188	0.0035220807
2,1	0.0065185901	0.0144927536	0.0016255757
2,2	0.0077257364	0.0072463768	0.0032511514
2,3	0.0079671656	0.0036231884	0.0024383636

```
$Hujan
Call: density.default(x = x, na.rm = TRUE)
Data: x (4142 obs.)      Bandwidth 'bw' = 0.8888

      x          y
Min. : 57.33  Min. :2.460e-06
1st Qu.: 68.17 1st Qu.:6.736e-04
Median : 79.00 Median :1.203e-02
Mean   : 79.00 Mean  :2.305e-02
3rd Qu.: 89.83 3rd Qu.:4.254e-02
Max.  :100.67  Max. :6.965e-02

$`Hujan Ekstrem`
Call: density.default(x = x, na.rm = TRUE)
Data: x (276 obs.)      Bandwidth 'bw' = 1.408

      x          y
Min. : 69.78  Min. :1.162e-05
1st Qu.: 77.89 1st Qu.:2.451e-03
Median : 86.00 Median :2.691e-02
Mean   : 86.00 Mean  :3.079e-02
3rd Qu.: 94.11 3rd Qu.:5.684e-02
Max.  :102.22  Max. :7.244e-02

$`Tidak Hujan`
Call: density.default(x = x, na.rm = TRUE)
Data: x (3691 obs.)      Bandwidth 'bw' = 1.169

      x          y
Min. :41.49  Min. :1.050e-06
1st Qu.:55.50 1st Qu.:5.769e-04
Median :69.50  Median :7.778e-03
Mean   :69.50  Mean  :1.783e-02
3rd Qu.:83.50 3rd Qu.:3.328e-02
Max.  :97.51  Max. :6.415e-02
```

b. Hasil Matriks Data *Training*

```
$Matrix_Uji
pred22      Hujan Hujan Ekstrem Tidak Hujan
  Hujan        3380      253       726
  Hujan Ekstrem    12       8        3
  Tidak Hujan     750      15      2962

$Akurasi
[1] 0.7830805

$Total_data
[1] 8109
```

c. Hasil Matriks Data *Testing*

```
$Matrix_Uji
```

		Hujan	Hujan	Ekstrem	Tidak Hujan	Hujan
		Hujan	815	60	206	
		Hujan Ekstrem	2	0	0	
		Tidak Hujan	172	8	764	

```
$Akurasi
```

```
[1] 0.7789837
```

```
$Total_data
```

```
[1] 2027
```

Lampiran 4. Syntax SVM

```
#----- SVM NEW -----#
library(e1071)
library(rpart)
library(MASS)

dataku=read.delim("clipboard")
nrow(dataku)
sampelsize=round(0.8*nrow(dataku))
set.seed(12345)
index = sample( seq_len ( nrow ( dataku ) ), size = samplesize )
training = dataku[ index, ]
View(training)
B=nrow(training)
B
testing=dataku [-index,]

#---- Mencari Parameter Terbaik Kernel RBF ----#
svm_tune=tune(svm,status ~., data = training,
               type="C-classification",
               kernel="radial",ranges =
list(cost=c(1,5,10,50,100),gamma=c(1,2,3,4,5,6,)))
summary(svm_tune)

#---- Mencari Parameter Terbaik Kernel Polynomial ----#
svm_tune=tune(svm,status ~., data = training,
               type="C-classification",
               kernel="polynomial",ranges = list(cost=c(1,5,10,50,100),d=c(1,2)))
summary(svm_tune)
```

```

#----- Model SVM RBF -----#
hujan.svm <- svm(status ~., data = training,
                    type="C-classification",
                    kernel="radial",
                    cost=1,gamma=1)
summary(hujan.svm)

#----- Model SVM Polynomial -----#
hujan.svm <- svm(status ~., data = training,
                    type="C-classification",
                    kernel="polynomial",
                    cost=5,d=1)
summary(hujan.svm)

# ----- Data Training ----- #
x <- training
pred2 <- predict(hujan.svm,x)
cm2 <- table(pred2,x$status)
accuracy2 <- sum(cm2[1,1]+cm2[2,2]+cm2[3,3])/sum(cm2)
accuracy2
list(Matrix_Uji=cm2,Akurasi=accuracy2,Total_data=sum(cm2))

# ----- Data Testing ----- #
xsvmtest <- testing
predsvmtest <- predict(hujan.svm,xsvmtest)
cmsvmtest <- table(predsvmtest,xsvmtest$status)
accuracysvmtest <-
sum(cmsvmtest[1,1]+cmsvmtest[2,2]+cmsvmtest[3,3])/sum(cmsvmtest)
accuracysvmtest
list(Matrix_Uji=cmsvmtest,Akurasi=accuracysvmtest,Total_data=sum(cmsvmtest))
)

```

Lampiran 5. Output SVM dengan Kernel RBF

a. Hasil Tuning

```

Parameter tuning of 'svm':
- sampling method: 10-fold cross validation
- best parameters:
  cost gamma
    1      1
- best performance: 0.2047136

- best performance: 0.2047136

- Detailed performance results:
  cost gamma   error dispersion
1     1       1 0.2047136 0.01302487
2     5       1 0.2066872 0.01583575
3    10       1 0.2093998 0.01455538
4    50       1 0.2132233 0.01376498
5   100       1 0.2184030 0.01497143
6     1       2 0.2071804 0.01428447
7     5       2 0.2114970 0.01514560
8    10       2 0.2166769 0.01617789
9    50       2 0.2383795 0.01521457
10   100      2 0.2484919 0.01467875
11   1       3 0.2071809 0.01553957
12   5       3 0.2214859 0.01455589
13   10      3 0.2333242 0.01651466
14   50      3 0.2631662 0.01499453
15   100     3 0.2757436 0.01565302
16   1       4 0.2097709 0.01681392
17   5       4 0.2339410 0.01542175
18   10      4 0.2454093 0.01548983
19   50      4 0.2790754 0.01816318
20   100     4 0.2947366 0.01441191
21   1       5 0.2129768 0.01660129
22   5       5 0.2404769 0.01797153
23   10      5 0.2539174 0.01464766
24   50      5 0.2946137 0.01523795
25   100     5 0.3064534 0.01552700
26   1       6 0.2171687 0.01419642
27   5       6 0.2475057 0.01527213
28   10      6 0.2627966 0.01353708
29   50      6 0.3017665 0.01283247
30   100     6 0.3122487 0.01353887

```

```
> svm_tune$best.model
Call:
best.tune(method = "svm", train.x = status ~ ., data = training, ranges = list(cost = c(1, 5, 10, 50,
100), gamma = c(1, 2, 3, 4, 5, 6)), type = "C-classification", kernel = "radial")

Parameters:
  SVM-Type: C-classification
  SVM-Kernel: radial
    cost: 1
    gamma: 1

Number of Support Vectors: 4145
```

b. Hasil Model

```
Call:
svm(formula = status ~ ., data = training, type = "C-classification", kernel = "radial",
cost = 1, gamma = 1)

Parameters:
  SVM-Type: C-classification
  SVM-Kernel: radial
    cost: 1
    gamma: 1

Number of Support Vectors: 4145
```

c. Hasil Matriks Data Training

```
$Matrix_Uji

pred2      Hujan Hujan Ekstrem Tidak Hujan
  Hujan      3628      267      783
  Hujan Ekstrem   0      0      0
  Tidak Hujan   514      9      2908

$Akurasi
[1] 0.806018

$Total_data
[1] 8109
```

d. Hasil Matriks Data Testing

```
$Matrix_Uji

predsvmtest      Hujan Hujan Ekstrem Tidak Hujan
  Hujan        868      66      209
  Hujan Ekstrem   0      0      0
  Tidak Hujan   121      2      761

$Akurasi
[1] 0.8036507

$Total_data
[1] 2027
```

Lampiran 6. Output SVM dengan Kernel Polynomial

a. Hasil Tuning

```

Call:
best.tune(method = svm, train.x = status ~ ., data = training, ranges = list(cost = c(1,
5, 10, 50, 100), d = c(1, 2)), type = "c-classification", kernel = "polynomial")

Parameters:
  SVM-Type: c-classification
  SVM-Kernel: polynomial
    cost: 5
   degree: 1
   gamma: 0.25
  coef.0: 0

Number of Support Vectors: 3913

Parameter tuning of 'svm':
- sampling method: 10-fold cross validation

- best parameters:
  cost d
      5 1

- best performance: 0.2106278

- Detailed performance results:
  cost d     error dispersion
1     1 1 0.2106279 0.01219221
2     5 1 0.2106278 0.01225891
3    10 1 0.2107512 0.01204768
4    50 1 0.2107512 0.01204768
5   100 1 0.2106279 0.01201074
6     1 2 0.4175613 0.01187991
7     5 2 0.4178081 0.01134073
8    10 2 0.4181780 0.01125885
9    50 2 0.4180545 0.01120115
10   100 2 0.4179312 0.01091846

```

b. Hasil Model

```

Call:
svm(formula = status ~ ., data = training, type = "c-classification", kernel = "polynomial",
cost = 5, d = 1)

Parameters:
  SVM-Type: c-classification
  SVM-Kernel: polynomial
    cost: 5
   degree: 1
   gamma: 0.25
  coef.0: 0

Number of Support Vectors: 3913

```

c. Hasil Matriks Data Training

```
$Matrix_Uji  
  
pred2          Hujan Hujan Ekstrem Tidak Hujan  
  Hujan        3525      267      809  
  Hujan Ekstrem  0        0        0  
  Tidak Hujan   617      9      2882  
  
$Akurasi  
[1] 0.7901098  
  
$Total_data  
[1] 8109
```

d. Hasil Matriks Data Testing

```
$Matrix_Uji  
  
predsvmtest    Hujan Hujan Ekstrem Tidak Hujan  
  Hujan        853      67      213  
  Hujan Ekstrem  0        0        0  
  Tidak Hujan   136      1      757  
  
$Akurasi  
[1] 0.7942773  
  
$Total_data  
[1] 2027
```