

**ANALISIS PERBANDINGAN METODE NAIVE BAYES CLASSIFIER  
DAN SUPPORT VECTOR MACHINE DALAM KLASIFIKASI  
CURAH HUJAN**

**SKRIPSI**



disusun oleh

**Ageng Bimantara**

**16.11.0476**

**PROGRAM SARJANA  
PROGRAM STUDI INFORMATIKA  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA**

**2020**

**ANALISIS PERBANDINGAN METODE NAIVE BAYES CLASSIFIER  
DAN SUPPORT VECTOR MACHINE DALAM KLASIFIKASI  
CURAH HUJAN**

**SKRIPSI**

untuk memenuhi sebagian persyaratan  
mencapai gelar Sarjana  
pada Program Studi Informatika



disusun oleh

**Ageng Bimantara**

**16.11.0476**

**PROGRAM SARJANA  
PROGRAM STUDI INFORMATIKA  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA**

**2020**

**PERSETUJUAN**

**SKRIPSI**

**ANALISIS PERBANDINGAN METODE NAIVE BAYES CLASSIFIER  
DAN SUPPORT VECTOR MACHINE DALAM KLASIFIKASI  
CURAH HUJAN**

yang dipersiapkan dan disusun oleh

**Ageng Bimantara**

**16.11.0476**

telah disetujui oleh Dosen Pembimbing Skripsi  
pada tanggal 8 Oktober 2019

**Dosen Pembimbing,**

**Ike Verawati, M.Kom.**

**NIK.190302237**

## PENGESAHAN

### SKRIPSI

#### ANALISIS PERBANDINGAN METODE NAIVE BAYES CLASSIFIER DAN SUPPORT VECTOR MACHINE DALAM KLASIFIKASI CURAH HUJAN

yang dipersiapkan dan disusun oleh

**Ageng Bimantara**

**16.11.0476**

telah dipertahankan di depan Dewan Penguji  
pada tanggal 15 Januari 2020

#### Susunan Dewan Penguji

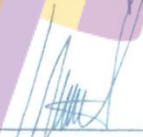
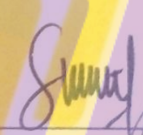
**Nama Penguji**

**Dina Maulina, M.Kom.**  
**NIK. 190302250**

**Sumarni Adi, S.Kom., M.Cs.**  
**NIK. 190302256**

**Ike Verawati, M.Kom.**  
**NIK. 190302237**

**Tanda Tangan**



Skripsi ini telah diterima sebagai salah satu persyaratan  
untuk memperoleh gelar Sarjana Komputer  
Tanggal 15 Januari 2020



**DEKAN FAKULTAS ILMU KOMPUTER**

**Krisnawati, S.Si, M.T.**  
**NIK. 190302038**

## PERNYATAAN

Saya yang bertanda tangan dibawah ini menyatakan bahwa, skripsi ini merupakan karya saya sendiri (ASLI), dan isi dalam skripsi ini tidak terdapat karya yang pernah diajukan oleh orang lain untuk memperoleh gelar akademis di suatu institusi pendidikan tinggi manapun, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis dan/atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Segala sesuatu yang terkait dengan naskah dan karya yang telah dibuat adalah menjadi tanggung jawab saya pribadi.

Yogyakarta, 20 Januari 2020



Ageng Bimantara

NIM. 16.11.0476

## MOTTO

*“Maka sesungguhnya bersama kesulitan ada kemudahan, Sesungguhnya bersama kesulitan ada kemudahan”.*

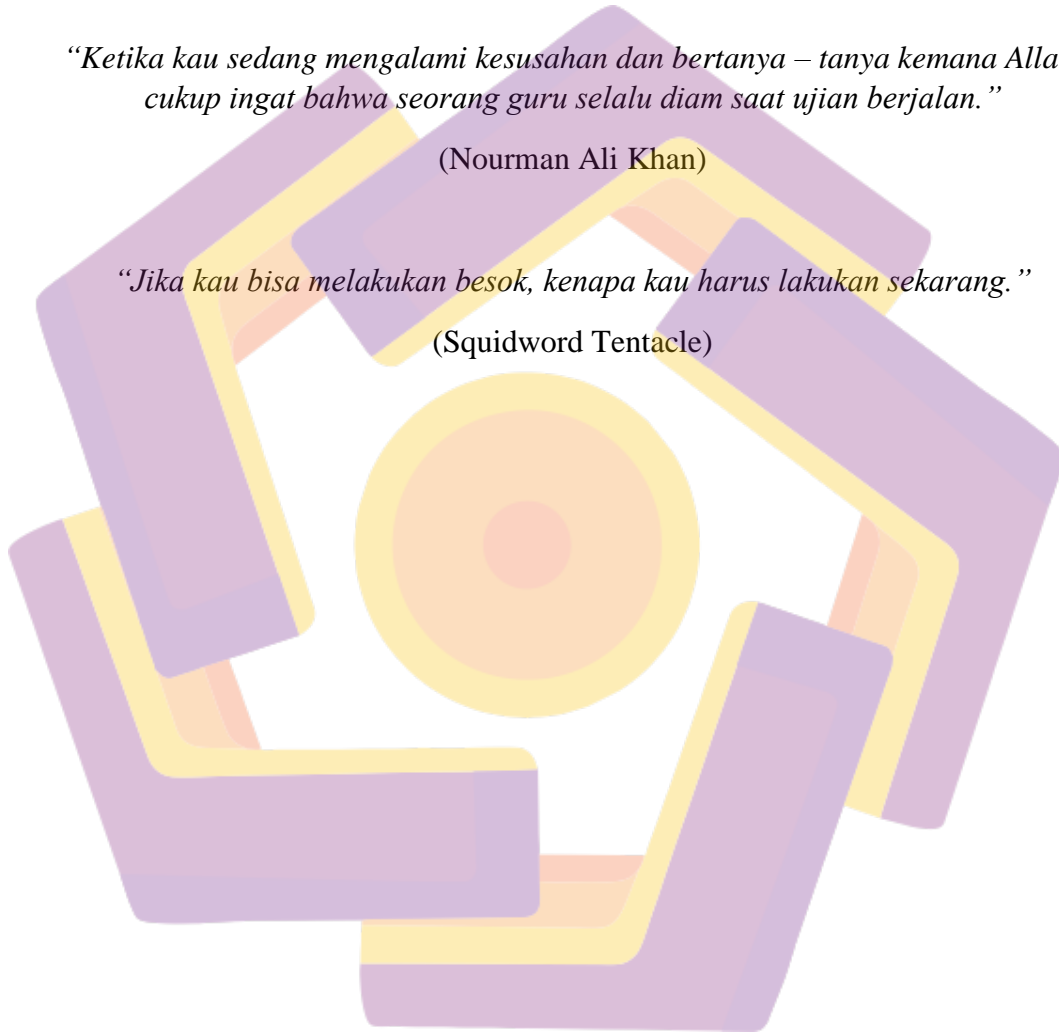
(Q.S Al – Insyirah : 5 – 6)

*“Ketika kau sedang mengalami kesusahan dan bertanya – tanya kemana Allah, cukup ingat bahwa seorang guru selalu diam saat ujian berjalan.”*

(Nourman Ali Khan)

*“Jika kau bisa melakukan besok, kenapa kau harus lakukan sekarang.”*

(Squidword Tentacle)



## PERSEMBAHAN

Dengan segala puji syukur kepada Allah SWT yang telah melimpahkan rahmat dan karunia – Nya sehingga skripsi ini dapat terselesaikan dengan baik.

Skripsi ini saya persembahkan untuk :

1. Kedua orang tua tercinta Bapak Parjimin dan Ibu Iswanti, yang selalu mendoakan dan memberikan dukungan.
2. Kakak kandung Anis Sholikhah yang selalu memberikan dukungan
3. Ibu Ike Verawati sebagai dosen pembimbing yang telah membantu selama skripsi berlangsung. Terima kasih atas saran, kritik, dan bantuannya
4. Azka Kurniawan Al Fathu dan Timur Dian Radha Sejati yang telah menjadi teman seperjuangan saat pendadaran.
5. Teman SMA, Hazazy yang telah menyediakan tempat kontrakkannya untuk menyelesaikan skripsi.
6. Keluarga besar 16 Informatika 08 terimakasih telah menjadi bagian dari keluarga dan memberi dukungan.
7. Seluruh pihak yang tidak dapat saya sebutkan satu per satu, terimakasih atas bantuan dan do'anya sehingga skripsi ini dapat terselesaikan.

## KATA PENGANTAR

*Alhamdulillah*, puji syukur penulis panjatkan kepada Allah SWT yang telah melimpahkan karunia-Nya sehingga penulis dapat menyelesaikan skripsi tepat waktu dengan judul **"ANALISIS PERBANDINGAN METODE NAIVE BAYES CLASSIFIER DAN SUPPORT VECTOR MACHINE DALAM KLASIFIKASI CURAH HUJAN"**. Skripsi ini disusun untuk memenuhi salah satu persyaratan dalam menyelesaikan Program Studi Strata-1 Informatika di Universitas Amikom Yogyakarta. Rasa terima kasih penulis sampaikan kepada seluruh pihak yang telah membantu, membimbing dan mendukung, khususnya kepada :

1. Bapak Prof. Dr. M. Suyanto, MM. selaku Rektor Universitas Amikom Yogyakarta.
2. Ibu Ike Verawati, M.Kom. selaku Dosen Pembimbing.
3. Bapak dan ibu Dosen Universitas Amikom Yogyakarta.
4. Seluruh teman-teman Mahasiswa/i Universitas Amikom Yogyakarta.

Penulis juga memohon maaf apabila dalam pelaksanaan dan penyusunan laporan skripsi ini terdapat kesalahan.

Yogyakarta, 20 Januari 2020

Penulis



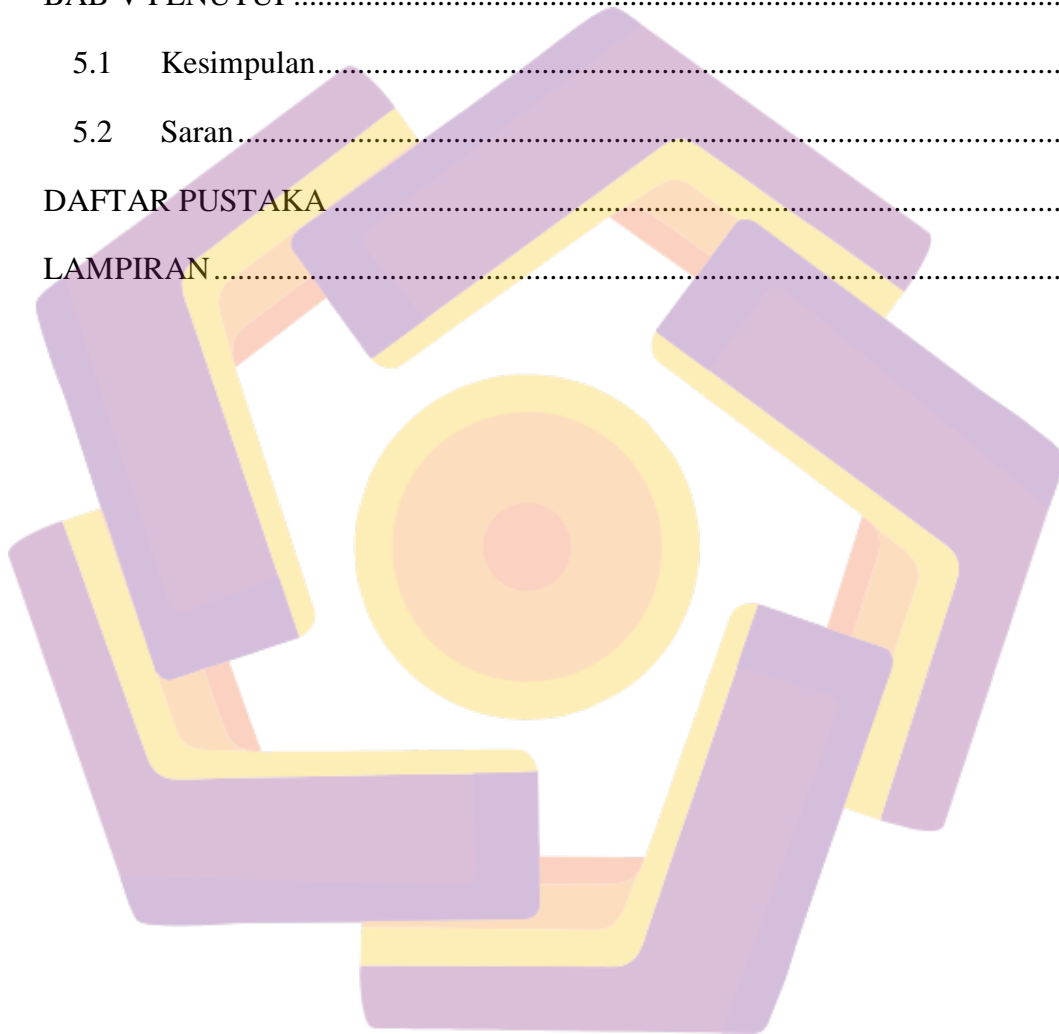
## DAFTAR ISI

JUDUL .....	i
PERSETUJUAN .....	i
PENGESAHAN .....	ii
PERNYATAAN.....	iii
MOTTO.....	iv
PERSEMBAHAN .....	v
KATA PENGANTAR .....	vi
DAFTAR ISI.....	vii
DAFTAR TABEL.....	xi
DAFTAR GAMBAR .....	xiii
INTISARI.....	xv
<i>ABSTRACT</i> .....	xv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah .....	2
1.3 Batasan Masalah.....	3
1.4 Tujuan Penelitian.....	3
1.5 Manfaat Penelitian.....	4
1.6 Metode Penelitian.....	4
1.61 Metode pengumpulan Data.....	4
1.62 Metode Testing .....	5
1.7 Sistematika Penulis.....	5

BAB II LANDASAN TEORI .....	6
2.1 Tinjauan Pustaka .....	6
2.2 Pengertian Curah Hujan .....	8
2.3 Variabel Berpengaruh dalam Curah Hujan .....	8
1. Rata – rata suhu .....	8
2. Rata – rata kelembaban .....	8
3. Lama Penyinaran .....	9
4. Rata – rata kecepatan angin .....	9
2.4 Data mining .....	9
2.4.1 <i>Association Rule Mining</i> .....	10
2.4.2 <i>Clustering</i> .....	10
2.4.3 Klasifikasi .....	10
2.5 Tahap klasifikasi.....	10
2.5.1 <i>Data Collection</i> .....	11
2.5.2 <i>Data Cleaning</i> .....	11
2.5.3 <i>Data Reduction</i> .....	11
2.6 <i>Data Training</i> dan <i>Data Testing</i> .....	11
2.7 Naive Bayes Classifier .....	12
2.8 <i>Support Vector Machine (SVM)</i> .....	14
2.9 <i>Non – linear Separable data</i> .....	15
2.10 <i>Confusion Matrix</i> .....	17
2.11 <i>Feature Selection</i> .....	19
2.11.1 <i>Information Gain</i> .....	19
2.11.2 <i>Chi – Square</i> .....	20
2.12 <i>K – Fold Cross Validation</i> .....	21

BAB III METODELOGI PENELITIAN .....	22
3.1 Alat dan Bahan .....	22
3.1.1 Alat Penelitian.....	22
3.1.2 Bahan Penelitian.....	22
3.2 Variabel Penelitian .....	23
3.2.1 Variabel Dependen.....	24
3.2.2 Variabel Independen .....	25
3.3 Alur Penelitian.....	26
3.3.1 Pengumpulan Data .....	27
3.3.2 <i>Data cleaning</i> .....	27
3.3.3 Proses Naive Bayes Classifier.....	27
3.3.4 Proses Support Vector Machine.....	27
3.3.5 NBC dengan <i>feature Selection</i> .....	28
3.3.6 SVM dengan <i>feature Selection</i> .....	28
BAB IV HASIL DAN PEMBAHASAN .....	29
4.1 Data Cleaning .....	29
4.2 Pembagian Data <i>Training</i> dan Data Testing .....	31
4.3 Analisis Naive Bayes Classifier .....	31
4.4 Analisis Klasifikasi <i>Support Vector Machine</i> .....	38
4.4.1 Analisis SVM dengan kernel RBF.....	40
4.4.2 Analisis SVM dengan Kernel Polynomial .....	44
4.5 Perbandingan NBC dan SVM .....	47
4.6 Analisis menggunakan <i>Features Selection</i> .....	47
4.7 Pembagian Data <i>Training</i> dan Data Testing .....	49
4.8 Analisis NBC dengan <i>Features Selection</i> .....	49

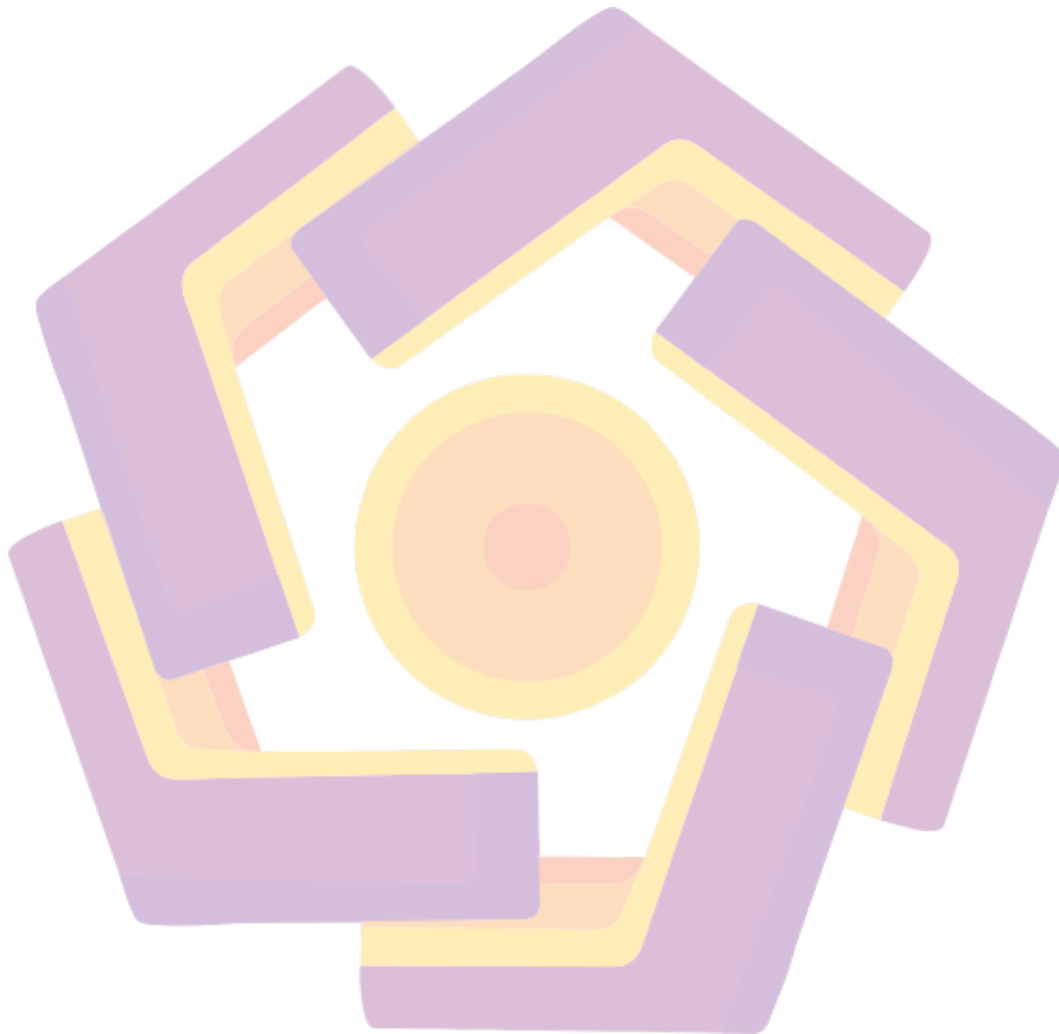
4.9	Support Vector Machine dengan <i>Feature Selection</i> .....	51
4.9.1	SVM Kernel RBF dengan <i>Feature Selection</i> .....	51
4.9.2	SVM Kernel Polynomial dengan <i>Feature Selection</i> .....	54
4.10	Perbandingan Hasil Akurasi .....	57
BAB V PENUTUP.....		59
5.1	Kesimpulan.....	59
5.2	Saran.....	60
DAFTAR PUSTAKA .....		61
LAMPIRAN.....		65



## DAFTAR TABEL

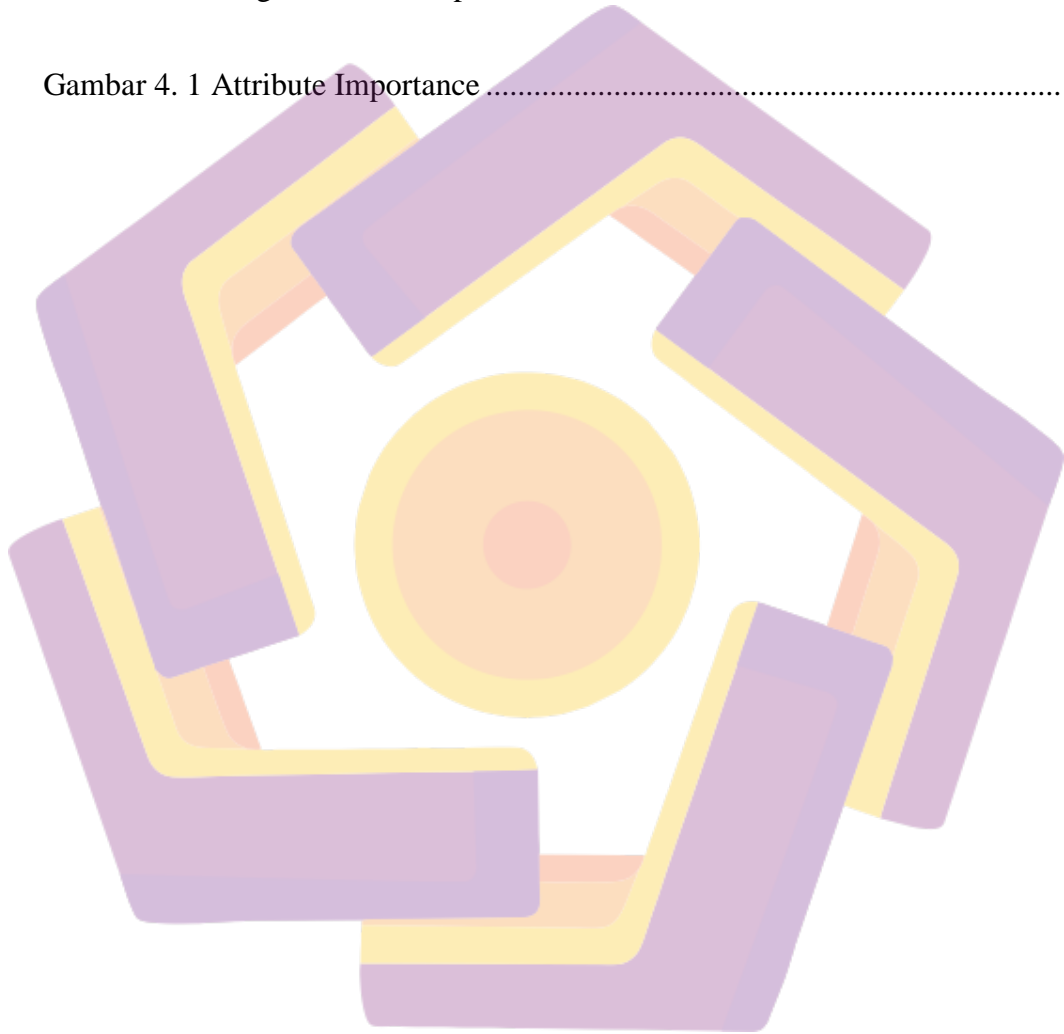
Tabel 2. 1 Confusion Matrix .....	17
Tabel 3. 1 Spesifikasi Perangkat Keras dan Perangkat Lunak.....	22
Tabel 3. 2 Sampel Data .....	23
Tabel 3. 3 Variable Penelitian.....	24
Tabel 3. 4 Kategori Variabel Dependen.....	24
Tabel 3. 5 Variabel Independen .....	25
Tabel 4. 1 Data Cuaca Harian Sebelum Data Cleaning .....	29
Tabel 4. 2 Data Cuaca Harian setelah Data Cleaning .....	30
Tabel 4. 3 Penggunaan Data Training dan Data Testing .....	31
Tabel 4. 4 Skala Curah Hujan .....	32
Tabel 4. 5 Priori Probability.....	35
Tabel 4. 6 Hasil Confusion Matrix NBC Data Training.....	35
Tabel 4. 7 Confusion Matrix Data Testing NBC .....	37
Tabel 4. 8 Skala Curah Hujan pada SVM.....	38
Tabel 4. 9 Akurasi SVM dengan Kernel RBF Data Training .....	40
Tabel 4. 10 Akurasi SVM dengan Kernel RBF Data Testing.....	41
Tabel 4. 11 Confusion Matrix SVM dengan Kernel RBF (C=1, Gamma = 1).....	42
Tabel 4. 12 Akurasi SVM dengan Kernel Polynomial Data Training .....	44
Tabel 4. 13 Akurasi SVM dengan Kernel Polynomial Data Testing.....	44
Tabel 4. 14 Confusion Matrix Kernel Polynomial (C=100, degree=1) .....	45
Tabel 4. 15 Hasil Perbandingan .....	47
Tabel 4. 16 Dataset yang Digunakan .....	48
Tabel 4. 17 Penggunaan Data Training dan Data Testing .....	49
Tabel 4. 18 Confusion Matrix NBC Data Training dengan Chi – Square .....	50
Tabel 4. 19 Confusion Matrix Data Testing NBC .....	51
Tabel 4. 20 Akurasi RBF dengan Feature Selection pada Data Training .....	51
Tabel 4. 21 Akurasi SVM Kernel RBF Data Testing .....	52
Tabel 4. 22 Confusion Matrix SVM Data Training dengan Chi – Square.....	53
Tabel 4. 23 Akurasi Kernel Polynomial Data Training dengan Chi – Square.....	54
Tabel 4. 24 Akurasi Kernel Polynomial Data Testing dengan Chi – Square.....	55

Tabel 4. 25 Confusion Matrix SVM data Training dengan Chi – Square..... 56  
Tabel 4. 26 Hasil Perbandingan Tanpa Feature Selection ..... 58  
Tabel 4. 27 Hasil Perbandingan Dengan Feature Selection..... 58



## DAFTAR GAMBAR

Gambar 2. 1 Hyperplane .....	15
Gambar 3. 1 Diagram Alir Tahapan Penelitian.....	26
Gambar 4. 1 Attribute Importance .....	48



## INTISARI

Curah hujan merupakan salah satu elemen pada iklim. Intesitas curah hujan yang tinggi dapat mengakibatkan banjir. Prakiraan cuaca menjadi salah satu solusi untuk mencegah banjir.

Pada penelitian ini menggunakan metode Naive Bayes Classifier (NBC) dan Support Vector Machine (SVM) dengan menggunakan *feature selection Chi – square* untuk meningkatkan akurasi dan mengatasi kekurangan dari kedua metode. Data yang digunakan berasal dari BMKG dari tahun 2016 sampai tahun 2018 pada provinsi Daerah Istimewa Yogyakarta.

Hasil dari penelitian menunjukkan NBC menghasilkan akurasi sebesar 79,019%, SVM pada kernel RBF  $C=1$   $\Gamma = 1$  menghasilkan akurasi sebesar 98,97%, SVM pada Kernel Polynomial  $C=100$  Degree = 1 menghasilkan akurasi sebesar 74,45%. Dengan menggunakan *feature selection* menunjukkan NBC menghasilkan akurasi sebesar 78,79%, SVM pada kernel RBF  $C=1$   $\Gamma = 1$  menghasilkan akurasi sebesar 99,77%, SVM pada Kernel Polynomial  $C=50$  Degree = 1 menghasilkan akurasi sebesar 77,53%.

**Kata Kunci :** NBC, SVM, *feature selection*, *chi – square*, Klasifikasi



## **ABSTRACT**

*Rainfall is one of element in the climate. The intensity of high rainfall can lead to flooding. Weather forecasting is one solution to prevent flooding. In this research using Naive Bayes Classifier (NBC) and Support Vector Machine method and use feature selection chi – square for improve accuracy.*

*This research used data in BMKG from 2016 until 2018 in Special Region of Yogyakarta.*

*The result of this research shown for NBC method has accuracy of 79,019%, SVM method with kernel RBF  $C=1$  and  $\Gamma=1$  has accuracy of 98,97%, and for SVM method with kernel Polynomial  $C=100$  Degree=1 has accuracy of 74,45%. Using feature selection NBC method has accuracy of 78,79%, SVM method with kernel RBF  $C=1$   $\Gamma=1$  has accuracy of 99,77%, and SVM method with kernel Polynomial  $C=500$  Degree=1 has accuracy 77,53%.*

**Key Word :** *NBC, SVM, feature selection, Chi – Square, Classification*