KLASIFIKASI CURAH HUJAN MENGGUNAKAN METODE ENSEMBLE SUBSET k-NEAREST NEIGHBOR

(Studi Kasus : Curah Hujan Kota Bogor Tahun 2014 – 2018)

TUGAS AKHIR

Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh Gelar Sarjana

Program Studi Statistika

Disusun Oleh:

N a m a : Meila Ika Pradipta

NIM : 15611077

PROGRAM STUDI STATISTIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS ISLAM INDONESIA

YOGYAKARTA

2020

HALAMAN PERSETUJUAN PEMBIMBING

TUGAS AKHIR

Judul	:	Klasifikasi Curah Hujan Menggunakan Metode Ensemble Subset k-Nearest Neighbor (Studi Kasus : Curah Hujan Kota Bogor Tahun 2014-2018)
Nama Mahasiswa	:	Meila Ika Pradipta
Nomor Mahasiswa	:	15611077

TUGAS AKHIR INI TELAH DIPERIKSA DAN DISETUJUI UNTUK DIUJIKAN

Yogyakarta, 27 Agustus 2020

Pembimbing

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

HALAMAN PENGESAHAN

TUGAS AKHIR

KLASIFIKASI CURAH HUJAN MENGGUNAKAN METODE ENSEMBLE SUBSET k-NEAREST NEIGHBOR

(Studi Kasus : Curah Hujan Kota Bogor Tahun 2014-2018)

Nama Mahasiswa Meila Ika Pradipta

Nomor Mahasiswa 15611077

TUGAS AKHIR INI TELAH DIUJIKAN

PADA TANGGAL 27 Agustus 2020

Nama Penguji	**Tanda Tangan**
1. Dr. Jaka Nugraha, S.Si., M.Si	.....
2. Muhammad Hasan Sidiq K, S.Si., M.Sc	.....
3. Ayundyah Kesumawati, M.Si.	.....

Mengetahui,

Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam

Prof. Riyanto, S.Pd., M.Si., Ph.D.

KATA PENGANTAR

Assalamualaikum Wr. Wb

Puji syukur penulis panjatkan kehadirat Allah SWT karena atas berkat, rahmat, kesehatan, kekuatan serta hidayah-Nya tugas akhir ini dapat berjalan dengan lancar.

Keberhasilan pembuatan tugas akhir ini tidak terlepas dari berbagai pihak yang membantu memberi semangat dan dukungan selama penyusunan tugas akhir ini. Pada kesempatan ini, penulis mengucapkan terima kasih kepada :

1. Bapak Prof. Riyanto S.Pd., M.Si., Ph.D selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Indonesia, Yogyakarta.
2. Bapak Dr. Edy Widodo S.Si., M.Si selaku Ketua Jurusan Statistika beserta seluruh jajarannya.
3. Ibu Ayundyah Kesumawati M.Si selaku Dosen Pembimbing yang telah sabar membimbing.
4. Seluruh dosen Statistika Universitas Islam Indonesia yang telah memberikan ilmu selama delapan semester.
5. Bapak, Ibu, Adik dan keluarga besar yang senantiasa mendoakan yang terbaik.
6. Paklik Sartono sekeluarga yang telah memotivasi dan memberi support serta telah menjadi keluarga kedua saya selama di Jogja.
7. Sahabat seperjuangan yang saya sayangi yakni Anggi Prabaningrum, Buyut Hiska R, Rahmatia G. Ali, Hani‘atul Maghfuroh, Zulfa Aulia Khusna, Aulia Zulfaniar dan Nurul Hitayuwana yang telah memberi semangat, motivasi serta sabar mendengarkan keluh kesah saya.
8. Teman-teman seperbimbingan yakni Listari, Nita Tri Anggraini, Hani‘atul Maghfuroh, Nida Nurhafidzah, Hadyanti Utami, Meymunah, Amalia Dwi dan Rhesa Mahardhika yang berjuang bersama untuk menyelesaikan tugas akhir ini.
9. Teman-teman Statistika 2015 yang sudah banyak memberikan semangat dan motivasi dalam memulai dan mengakhiri tugas akhir ini.

10. Semua pihak yang tidak dapat penulis sebutkan satu persatu.

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. Akhir kata, semoga Allah SWT selalu melimpahkan rahmat serta hidayah-Nya kepada kita semua. Aamiin aamiin ya robbal _alamin.

Wassalamualaikum, Wr. Wb

Yogyakarta, 27 Agustus 2020

(Meila Ika Pradipta)

HALAMAN PERNYATAAN KEASLIAN TUGAS AKHIR

Menyatakan bahwa seluruh komponen dan isi dalam tugas akhir ini adalah hasil karya saya sendiri. Apabila dikemudian hari terbukti ada beberapa bagian dari karya ini adalah bukan hasil karya sendiri, tugas akhir yang diajukan sebagai hasil karya sendiri ini siap ditarik kembali dan siap menanggung resiko dan konsekuensi apapun.

Demikian surat pernyataan ini dibuat, semoga dapat dipergunakan sebagaimana mestinya.

Yogyakarta, 27 Agustus 2020

(Meila Ika Pradipta)

HALAMAN PERSEMBAHAN

Skripsi ini ananda persembahkan untuk :

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

Adik peneliti yang selalu memberikan doa, semangat, motivasi serta selalu bersedia mendengarkan keluh kesah peneliti.

HALAMAN MOTO

Man Jadda Wajadda

(Barangsiapa yang bersungguh-sungguh, pasti akan berhasil)

Success needs a process

(Sukses membutuhkan suatu proses)

There is no limit of struggling

(Tidak ada batasan untuk berjuang)

Stop dreaming, start doing!

(Berhenti bermimpi, saatnya beraksi!)

Memulai dengan penuh keyakinan
Menjalankan dengan penuh keikhlasan
Menyelesaikan dengan penuh kebahagiaan

KLASIFIKASI CURAH HUJAN MENGGUNAKAN METODE ENSEMBLE SUBSET K-NEAREST NEIGHBOR
(Studi Kasus : Curah Hujan Kota Bogor Tahun 2014-2018)

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

INTISARI

Curah hujan merupakan salah satu komponen dalam iklim. Klasifikasi curah hujan menjadi salah satu masalah yang cukup menarik. Besarnya curah hujan tidak dapat ditentukan secara pasti namun dapat diperkirakan. Metode kNN merupakan sebuah pendekatan untuk klasifikasi non-parametrik yang cukup efisien. Berbagai upaya telah dilakukan untuk meningkatkan kinerja tetangga terdekat yang diklasifikasikan dengan teknik ensemble. Oleh karena itu, diperlukan suatu teknik khusus untuk melakukan klasifikasi terhadap data curah hujan yang dinamakan Ensemble Subset K-Nearest Neighbor. Tujuan penelitian ini adalah untuk mengetahui klasifikasi curah hujan. Data yang digunakan merupakan data sekunder yang bersumber dari website resmi BMKG tahun 2014-2018. Hasil dari penelitian ini menunjukkan bahwa ketika nilai k sebesar 3, maka nilai error yang didapatkan 0.260223 dan ketepatan klasifikasi yang dihasilkan sebesar 0.739777. Ketika nilai k sebesar 4, maka nilai error yang didapatkan 0.2695167 dan ketepatan klasifikasi yang dihasilkan sebesar 0.730483. Ketika nilai k sebesar 5, maka nilai error yang didapatkan 0.2695167 dan ketepatan klasifikasi yang dihasilkan sebesar 0.730483. Ketika nilai k sebesar 6, maka nilai error yang didapatkan sebesar 0.260223 dan ketepatan klasifikasi yang dihasilkan sebesar 0.739777.

Kata kunci: Curah Hujan, ESKNN, Error, Ketepatan klasifikasi.

RAINFALL CLASSIFICATION USING K-NEAREST NEIGHBOR ENSEMBLE SUBSET METHOD

(Case Study : Bogor Rainfall in 2014 to 2018)

Meila Ika Pradipta

Department Statistics, Faculty Mathematic and Natural Sciences

Islamic University of Indonesia

ABSTRACT

Rainfall is one component in the climate. Classification of rainfall is one of the problems that is quite interesting. The amount of rainfall cannot be determined exactly but it can be estimated. The kNN method is an approach for non-parametric classification which is quite efficient. Various attempts have been made to improve the performance of the nearest neighbors who are classified with ensemble techniques. Therefore, a special technique is needed to classify the rainfall data called the Ensemble Subset K-Nearest Neighbor. The purpose of this study was to determine the classification of rainfall. The data used are secondary data sourced from the 2014-2018 BMKG official website. The results of this study indicate that when the k value is 3, the error value obtained is 0.260223 and the resulting classification accuracy is 0.739777. When the value of k is 4, the error value obtained is 0.2695167 and the resulting classification accuracy is 0.730483. When the value of k is 5, the error value obtained is 0.2695167 and the resulting classification accuracy is 0.730483. When the value of k is 6, the error value obtained is 0.260223 and the resulting classification accuracy is 0.739777.

Keywords: Rainfall, ESKNN, Error, Classification accuracy.

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

BAB I
PENDAHULUAN

1.1 Latar Belakang Masalah

Iklim merupakan salah satu hal yang cukup krusial di negara ini. Terutama Indonesia yang beriklim tropis. Iklim sendiri mempunyai pengaruh yang sangat besar bagi kelangsungan hidup manusia pada beberapa sektor diantaranya pertanian, pariwisata, transportasi, dan lain-lain. Iklim merupakan kondisi rata-rata cuaca dalam waktu yang relatif panjang. Cuaca dapat berubah seiring perubahan waktu. Salah satu faktor yang penting dalam perubahan cuaca yakni curah hujan.

Curah hujan merupakan salah satu komponen dalam iklim. Curah hujan juga merupakan hal yang dapat berpengaruh pada cuaca. Menurut BMKG, curah hujan adalah ketebalan air hujan yang terkumpul dalam luasan 1 m^2. Terdapat beberapa faktor yang mempengaruhi curah hujan diantaranya suhu, kelembaban, lama penyinaran, dan arah angin.

Pada beberapa waktu silam, prediksi curah hujan menjadi salah satu masalah yang cukup menarik. Besarnya curah hujan tidak dapat ditentukan secara pasti, namun dapat diprediksi atau diperkirakan. Untuk menghasilkan prediksi yang akurat, diperlukan usaha dan kerja keras oleh para ahli Klimatologi yang telah menjadikan masalah tersebut menjadi sebuah tantangan besar untuk dipecahkan.

Di Indonesia sendiri, Kota Bogor bukan merupakan kota yang menduduki peringkat sebagai kota yang memiliki curah hujan yang tinggi. Namun dikarenakan Kota Bogor memiliki julukan sebagai Kota Hujan yang tercatat bahwa hujan turun tidak mengenal waktu. Bahkan musim hujan atau kemarau pun akan tetap turun hujan. Maka dari itu, penelitian ini dilakukan untuk menghasilkan prediksi curah hujan di Kota Bogor yang akan datang.

Secara geografis, kota Bogor sendiri berada pada ketinggian minimum 190 m dan maksimum 330 m dari permukaan laut. Udara paling sejuk di Bogor mencapai 21.8 derajat celcius. Sehingga membuat kawasannya sering dilanda

hujan orografi (hujan yang terjadi di daerah pegunungan). (Bogor, 2016). Kota Bogor sendiri memiliki intensitas curah hujan yang cukup tinggi yakni di atas 100 mm per hari. Menurut salah satu petugas di Stasiun Klimatologi Dramaga yakni pak Budi, bahkan ketika puncak musim kemarau melanda Indonesia, Bogor tetap sering diguyur hujan. Selain itu, curah hujan di Bogor dapat dikatakan cukup ekstrim dikarenakan hujan diikuti dengan petir dan angin yang cukup kencang. Oleh karena itu, diperlukan suatu teknik khusus untuk melakukan klasifikasi terhadap data curah hujan.

Salah satu metode paling sederhana dan tertua untuk klasifikasi adalah klasifikasi k tetangga terdekat (kNN). Metode ini mengklasifikasikan pengamatan yang tidak diketahui ke kelas mayoritas di antara pengamatan tetangga terdekatnya, yang diukur dengan metrik jarak, dalam data pelatihan (Cover T, 1967). Metode *K-Nearest Neighbor* (kNN) merupakan salah satu algoritma *Machine Learning* (ML) yang dianggap sebagai suatu metode yang sederhana untuk diterapkan dalam analisis data dengan dimensi variabel yang banyak. (Alkhatib K, 2013). Walaupun metode ini sederhana, namun metode ini memiliki kelebihan dibandingkan metode lainnya, yaitu dapat menggeneralisasi himpunan data *training* yang relatif kecil. (Maimon, 2010). Terlepas dari kesederhanaannya, kNN memberikan hasil yang kompetitif dan dalam beberapa kasus bahkan mengungguli algoritma pembelajaran kompleks lainnya. Metode kNN merupakan sebuah pendekatan untuk klasifikasi non-parametrik yang cukup efisien. Namun, kNN dipengaruhi oleh fitur-fitur non-informatif dalam data, sering kali dengan data dimensi tinggi. Berbagai upaya telah dilakukan untuk meningkatkan kinerja tetangga terdekat yang diklasifikasikan dengan teknik *ensemble*. (Gul, 2016)

Berdasarkan penjelasan pada latar belakang diatas dan berdasar tinjauan pustaka yang telah di himpun, terdapat beberapa penelitian yang menggunakan data curah hujan. Selain itu alasan pemilihan metode yang terbilang baru dan dirasa cocok untuk mengatasi masalah curah hujan maka dari itu penulis berinisiatif mengangkat judul ‒Klasifikasi Data Curah Hujan Menggunakan Ensemble Subset K-Nearest Neighbor‖. Diharapkan penelitian mengenai klasifikasi prediksi curah hujan ini nantinya dapat mengantisipasi perubahan

cuaca yang tidak menentu di Kota Bogor serta dapat menjadi bahan pertimbangan bagi pembuat keputusan dan pihak terkait.

1.2 Rumusan Masalah

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

1. Bagaimana kondisi secara umum curah hujan di Kota Bogor pada tahun 2014 - 2018?
2. Bagaimana hasil klasifikasi curah hujan di Kota Bogor dengan menggunakan Ensemble Subset K-Nearest Neighbor?

1.3 Batasan Masalah

Adapun batasan masalah pada penelitian ini adalah sebagai berikut:

1. Penelitian ini menggunakan data yang diambil dari *website* BMKG dari tahun 2014 sampai tahun 2018 di Kota Bogor.
2. Alat analisis yang digunakan untuk memprediksi nilai curah hujan adalah deskriptif dan Ensemble Subset K-Nearest Neighbor.
3. *Software* yang digunakan adalah R 3.6.1
4. Untuk keseluruhan variabel yang digunakan merupakan data hasil rata-rata yang diperoleh dari *website* BMKG yaitu www.bmkg.go.id

1.4 Tujuan Penelitian

Penelitian ini memiliki tujuan sebagai berikut :

1. Untuk mengetahui kondisi secara umum curah hujan di Kota Bogor tahun 2014 – 2018.
2. Untuk mengetahui hasil klasifikasi curah hujan di Kota Bogor dengan menggunakan Ensemble Subset K-Nearest Neighbor.

1.5 Manfaat Penelitian

Manfaat dari penelitian ini adalah:

1. Dapat menjelaskan penggunaan metode *Ensemble Subset K-Nearest Neighbor*.

2. Dapat menjadi sarana informasi dan referensi bagi semua pihak yang tertarik dengan masalah yang dibahas dalam penelitian ini.

BAB II
TINJAUAN PUSTAKA

Setelah peneliti melakukan telaah pada beberapa penelitian terdahulu, terdapat beberapa penelitian yang memiliki keterkaitan dengan penelitian yang peneliti lakukan. Baik keterkaitan pada metode ataupun objek pada penelitian. Berikut beberapa keterkaitan penelitian terdahulu.

Penelitian yang dilakukan oleh Rakhmalia pada tahun 2018. Penelitian tersebut membahas tentang perbandingan hasil metode *Naïve Bayes Classifier* dan *Support Vector Machine* dalam klasifikasi curah hujan. Penelitian tersebut bertujuan untuk mendapatkan hasil perbandingan antara metode *Naïve Bayes Classifier* dan *Support Vector Machine*. Dalam penelitian tersebut, data yang digunakan merupakan data curah hujan harian di Provinsi Jawa Timur bulan Januari 2013 hingga Desember 2017 di seluruh stasiun BMKG di Jawa Timur. Variabel yang digunakan yaitu rata-rata suhu, rata-rata kelembaban, lama penyinaran, rata-rata kecepatan angin dan status curah hujan (yang telah di kategorikan). Hasil yang diperoleh yaitu 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.

Penelitian yang dilakukan oleh Puspaningrum pada tahun 2018. Penelitian tersebut membahas tentang perbandingan metode *Extreme Machine Learning* dan SARIMA/GARCH dalam peramalan data curah hujan. Penelitian tersebut bertujuan untuk mengetahui metode terbaik dari hasil peramalan menggunakan metode *Extreme Machine Learning* dan SARIMA/GARCH serta hasil dari peramalannya. Dalam penelitian tersebut, data yang digunakan merupakan data curah hujan per dasarian di Kabupaten Sleman dari tahun 1998-2017. Variabel yang digunakan yaitu hanya jumlah curah hujan saja. Hasil yang diperoleh yaitu rata-rata curah hujan dari keenam pos yang terdapat di Kabupaten Sleman pada

tahun 1998, 2016 dan 2017 tinggi. Berdasarkan nilai kesalahan peramalan didapatkan hasil bahwa metode ELM memiliki nilai *error* yang lebih kecil dibandingkan dengan metode SARIMA/GARCH untuk 5 pos stasiun. Sedangkan untuk satu pos sisanya, metode SARIMA/GARCH memiliki nilai kesalahan peramalan yang lebih kecil dibandingkan metode ELM. Namun, hasil peramalan dengan metode ELM belum dapat digunakan untuk melakukan peramalan data curah hujan dasarian karena terdapat hasil peramalan yang bernilai negatif dan terdapat beberapa nilai curah hujan yang memiliki selisih antara nilai peramalan dengan data asli yang cukup banyak.

Penelitian yang dilakukan oleh Humaini pada tahun 2015. Penelitian tersebut membahas tentang penggunaan metode *Extreme Machine Learning* untuk memprediksi kondisi cuaca di wilayah Malang. Penelitian tersebut bertujuan untuk mengetahui bagaimana model jaringan syaraf tiruan ELM untuk memprediksi kondisi cuaca di wilayah Malang. Dalam penelitian tersebut, data yang digunakan adalah data harian bulan Maret 2014 dan Agustus 2014 yang diambil dari instansi BMKG Karangploso Malang tahun 2014. Hasil dari penelitian tersebut diperoleh model optimal dari proses training jaringan saraf tiruan ELM untuk prediksi cuaca di Malang yang terdiri dari 4 unit *input* (kecepatan angin, suhu udara, kelembaban udara dan tekanan udara), 4 unit *hidden* pada 1 *hidden layer* dan 1 unit *output* (hujan atau tidak hujan). Sebagaimana hasil *output* jaringan saraf tiruan dengan menggunakan data *testing* menghasilkan 80% memenuhi kriteria data dan 20% tidak memenuhi kriteria data, dengan keterangan hujan (Maret 2014) dan tidak hujan (Agustus 2014). Sehingga pada prediksi data *testing* menggunakan model jaringan syaraf tiruan menghasilkan galat 20%.

Pada tahun 2016, Gul mempublikasikan jurnal yang berjudul ―*Ensemble of a Subset of kNN Classifiers*‖. Dalam jurnal tersebut, penulis menggabungkan beberapa klasifikasi, yang dikenal sebagai metode *ensemble*, dapat memberikan peningkatan substansial dalam kinerja prediksi algoritma pembelajaran terutama di hadapan fitur non-informatif dalam set data. Penulis mengusulkan *ensemble subset* dari klasifikasi kNN, ESkNN, untuk tugas klasifikasi dalam dua langkah.

Pertama, penulis memilih klasifikasi berdasarkan kinerja masing-masing menggunakan akurasi *out-of-sampel*. Klasifikasi yang dipilih kemudian digabungkan secara berurutan mulai dari model terbaik dan dinilai untuk kinerja kolektif pada set data validasi. Penulis menggunakan set data tanda *benchmark* dengan fitur asli dan beberapa tambahan yang tidak informatif untuk evaluasi metode penulis. Hasilnya dibandingkan dengan kNN biasa, kNN *bagged*, random kNN, metode subset fitur ganda, *random forest* dan *support vector machine*. Perbandingan eksperimental penulis pada masalah klasifikasi *benchmark* dan set data simulasi mengungkapkan bahwa *ensemble* yang diusulkan memberikan kinerja klasifikasi yang lebih baik daripada kNN biasa dan ensembelnya, dan membandingkan kinerja *random forest* dan *support vector machine*.

Pada tahun 2011, Thirey mempublikasikan jurnal yang berjudul ―*Increasing Accuracy Through Class Detection: Ensemble Creation Using Optimized Binary kNN Classifier*‖. Dalam jurnal tersebut, penulis menjelaskan bahwa *Ensemble Classifier* telah berhasil digunakan untuk meningkatkan tingkat akurasi klasifikasi yang mendasarinya. Melalui penggunaan klasifikasi teragregasi, sangat mungkin untuk mencapai tingkat kesalahan yang lebih rendah dalam klasifikasi daripada dengan menggunakan *instance classifier* tunggal. *Ensemble* paling sering digunakan dengan menggabungkan pohon keputusan atau jaringan saraf karena tingkat kesalahan yang lebih tinggi ketika digunakan secara individual. Dalam jurnal ini, penulis akan mempertimbangkan implementasi unik dari *ensemble classifier* yang menggunakan pengklasifikasi kNN. Setiap *classifier* dirancang untuk mendeteksi keanggotaan dalam kelas tertentu menggunakan proses seleksi subset terbaik untuk variabel. Ini dapat memberikan keragaman yang diperlukan untuk keberhasilan mengimplementasikan *ensemble*. Mekanisme agregasi untuk menentukan klasifikasi akhir dari *ensemble* disajikan dan diuji terhadap beberapa set data yang terkenal.

Pada tahun 2014, Hassanat mempublikasikan jurnal yang berjudul ―*Solving the Problem of the K Parameter in the KNN Classifier Using an Ensemble Learning Approach*‖ Dalam jurnal tersebut, penulis menyajikan solusi baru untuk memilih parameter *k* dalam algoritma *k*-tetangga terdekat (kNN),

solusinya tergantung pada gagasan pembelajaran *ensemble*, di mana *classifier* kNN yang lemah digunakan setiap kali dengan *k* yang berbeda, mulai dari satu ke akar kuadrat dari ukuran set pelatihan. Hasil dari pengklasifikasi lemah digabungkan menggunakan aturan jumlah tertimbang. Solusi yang diusulkan diuji dan dibandingkan dengan solusi lain menggunakan eksperimen dalam masalah kehidupan nyata. Hasil percobaan menunjukkan bahwa pengelompokan yang diusulkan mengungguli pengelompokan kNN tradisional yang menggunakan jumlah tetangga yang berbeda, kompetitif dengan pengelompokan lain, dan merupakan pengelompokan yang menjanjikan dengan potensi kuat untuk berbagai aplikasi.

Tabel 2.1 Rujukan Penelitian

Nama	**Variabel/Data**	**Metode**	**Kesimpulan**
Riza Indriani Rakhmalia (2018)	Rata-rata suhu, rata-rata kelembaban, lama penyinaran, rata-rata kecepatan angin dan status curah hujan (yang telah dikategorikan)	Naïve Bayes Classifier dan Support Vector Machine	Perbandingan nilai akurasi pada metode NBC dan SVM yang tertinggi adalah metode SVM menggunakan kernel RBF dengan C=1 dan Gamma=1. Hasil untuk akurasi data *testing* sebesar 80.365%
Laksmita Puspaningrum (2018)	Jumlah curah hujan	Extreme Machine Learning dan SARIMA/GARCH	Metode ELM memiliki nilai *error* yang lebih kecil dibandingkan metode SARIMA/GARCH untuk 5 pos stasiun. Sedangkan untuk satu pos sisanya, metode SARIMA/GARCH memiliki nilai kesalahan

Nama	Variabel/Data	Metode	Kesimpulan
			peramalan lebih kecil dibanding metode ELM
Qoid Humaini (2015)	Musim, tanggal, jam, kecepatan angin, suhu udara, kelembaban udara, tekanan udara	Extreme Machine Learning	Model optimal dari proses *training* jaringan syaraf tiruan ELM untuk prediksi cuaca di Malang terdiri dari 4 unit *input*, 4 unit *hidden* pada 1 *hidden layer* dan 1 unit *output*. Prediksi data *testing* menghasilkan galat 20%
Asma Gul (2016)	Set data *benchmark*, penulis menilai ESKNN menggunakan simulasi model. Berikut rincian kedua simulasi modelnya. Model pertama, variabel untuk kelas 1 berkorelasi dan dihasilkan dengan struktur varians kovarians yang bervariasi, sedangkan fitur yang menentukan kelas 2 adalah independen. Sebanyak 500 set data kelas biner independen	Ensemble of a Subset of kNN Classifiers	Hasilnya dibandingkan dengan kNN biasa, kNN bagged, random kNN, metode subset fitur ganda, *random forest* dan *support vector machine*. Perbandingan eksperimental penulis pada masalah klasifikasi *benchmark* dan set data simulasi mengungkapkan bahwa *ensemble* yang diusulkan memberikan kinerja klasifikasi yang lebih baik daripada kNN biasa dan ensembelnya, dan membandingkan kinerja *random forest* dan *support vector machine*

Nama	Variabel/Data	Metode	Kesimpulan
	dihasilkan, masing-masing dengan 20 fitur. Model simulasi kedua adalah model 4 dimensi yang berasal dari model yang diusulkan dalam Mease et al. (2007). Seperangkat 500 set data kelas biner independen dihasilkan, masing-masing terdiri dari 1000 pengamatan dan 4 fitur		
Benjamin Thirey (2011)	Data yang digunakan berasal dari UCI *Machine Learning Repository* dengan pengecualian dataset IRIS yang tersedia dalam perangkat lunak R (23, 24, 25). Selain itu juga menggunakan data *Low Resolution Spectrometer* (LRS) dan dataset ARRYTHMIA	Ensemble Creation Using Optimized Binary kNN Classifier	Tingkat akurasi dan jumlah variabel yang digunakan disajikan pada tabel yang terdapat pada jurnal tersebut. Dengan menggunakan pengklasifikasi individual, didapatkan tingkat klasifikasi yang tinggi. Dalam dataset LRS dan ARRYTHMIA, model dalam klasifikasi *ensemble* hanya menggunakan sekitar 5% dari variabel prediktor yang tersedia. Jumlah rata-rata variabel

Nama	Variabel/Data	Metode	Kesimpulan
			prediktor yng digunakan untuk klasifikasi secara signifikan lebih kecil dari itu
Ahmad Basser Hassanat (2014)	Data yang digunakan yaitu 28 set data yang berbeda untuk mewakili klasifikasi yang diambil dari UCI *Machine Learning Repository* yang terdapat pada tabel 1 dalam jurnal tersebut	Solving the Problem of the K Parameter in the KNN Classifier Using an Ensemble Learning Approach	Hasil eksperimen menggunakan berbagai set data menunjukkan keunggulan metode yang diusulkan menggunakan berbagai nilai *k*. Selain itu, kecepatan yang diusulkan menunjukkan bahwa metode (waktu linier) lebih baik daripada metode IINC yang merupakan waktu linier logaritmik
Meila Ika Pradipta (2020)	Tanggal, suhu rata-rata, kelembaban rata-rata, curah hujan, lama penyinaran, dan arah angin saat kecepatan maksimum	Ensemble Subset KNN	Hasil klasifikasi yang didapatkan telah dijelaskan di penelitian ini terdapat pada bab V Pembahasan. Untuk ketepatan klasifikasi dan nilai errornya didapatkan hasil ketika nilai k sebesar 3, maka nilai error yang didapatkan 0.260223 dan ketepatan klasifikasi yang dihasilkan sebesar 0.739777. Ketika nilai *k* sebesar 4, maka nilai error yang didapatkan

Nama	Variabel/Data	Metode	Kesimpulan
			0.2695167 dan ketepatan klasifikasi yang dihasilkan sebesar 0.730483. Ketika nilai k sebesar 5, maka nilai error yang didapatkan 0.2695167 dan ketepatan klasifikasi yang dihasilkan sebesar 0.730483. Ketika nilai k sebesar 6, maka nilai error yang didapatkan sebesar 0.260223 dan ketepatan klasifikasi yang dihasilkan sebesar 0.739777.

BAB III
LANDASAN TEORI

3.1 Curah Hujan

Curah hujan adalah jumlah air yang jatuh di permukaan tanah datar selama periode tertentu yang diukur dengan satuan tinggi milimeter (mm) di atas permukaan horizontal. Curah hujan 1 (satu) milimeter, artinya dalam luasan satu meter persegi pada termpat yang datar tertampung air setinggi satu milimeter atau tertampung air sebanyak satu liter. (Manalu, 2016). Berdasarkan penelitian yang dilakukan oleh (Fauziah, 2008), terdapat lima kriteria curah hujan sebagai berikut:

Tabel 3.1 Kriteria Curah Hujan

Status	Rentang Curah Hujan
Sangat Ringan	< 5 mm/hari
Ringan	5 – 20 mm/hari
Sedang	21 – 50 mm/hari
Lebat	51 – 100 mm/hari
Sangat Lebat	>100 mm/hari

3.2 Geografis Kota Bogor

Secara geografis, Kota ogor terletak di antara 106 4 T dan 6 26 LS. Kota Bogor mempunyai rata-rata ketinggian minimum 190 m dan maksimum 350 m dari permukaan laut. (Diskominfostandi Kota Bogor, 2016). Menurut *website* (Bogor, 2016), kondisi iklim di Kota ogor suhu rata-rata tiap bulan 26 C dengan suhu terendah 21. C dengan suhu tertinggi 30.4 C. Kelembaban udara 70 %, Curah hujan rata-rata setiap tahun sekitar 3.500 – 4000 mm dengan curah hujan terbesar pada bulan Desember dan Januari. Luas Wilayah Kota Bogor sebesar 11.850 Ha yang terdiri dari 6 kecamatan dan 68 kelurahan. Kemudian secara administratif, kota Bogor terdiri dari 6 wilayah kecamatan, 31 kelurahan dan 37 desa (lima diantaranya termasuk desa tertinggal yaitu desa Pamoyanan, Genteng, Balungbangjaya, Mekarwangi dan Sindangrasa), 210 dusun, 623 RW, 2.712 RT.

3.3 Variabel Berpengaruh dalam Curah Hujan

Berdasarkan penelitian yang dilakukan oleh (Rakhmalia, 2018) terdapat beberapa variabel yang berpengaruh terhadap curah hujan.

a. Rata-rata uhu C)

   Menurut MKG ― uhu 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 C).

b. Rata-rata Kelembaban (%)

   Kelembaban udara menurut MKG 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 kelembapan udara adalah higrometer.‖ Pada variabel independen rata-rata kelembaban ini data yang digunakan berskala numerik dalam bentuk persentase.

c. Lama Penyinaran (jam)

   Lama penyinaran artinya adalah lama penyinaran oleh 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 MKG‖. Data pada variabel lama penyinaran dalam satuan jam.

d. Arah Angin Saat Kecepatan Maksimum (deg)

   Arah angin adalah petunjuk pergerakan angin. Arah angin juga merupakan dari mana angin tersebut berhembus dan dinyatakan dengan sudut di dalam kompas. Data pada variabel arah angin dalam satuan deg.

3.4 Software R

R termasuk kelompok perangkat statistik *open source* yang tidak memerlukan lisensi/gratis, atau yang dikenal dengan *freeware*. Paket R memiliki fasilitas yang sangat banyak untuk analisis data statistik, mulai dari metode yang klasik sampai dengan yang modern, seperti pengolahan data frekuensi, korelasi, regresi, dan lain sebagainya.

Adapun kehebatan-kehebatan dari R adalah:

1. Efektif dalam pengelolaan data dan fasilitas penyimpanan. Ukuran *file* yang disimpan jauh lebih kecil dibanding perangkat lunak lainnya.
2. Lengkap dalam operator perhitungan *array*.
3. Lengkap dan terdiri dari koleksi alat statistik yang terintegrasi untuk analisis data, diantaranya, mulai statistik deskriptif, fungsi probabilitas, berbagai macam uji statistik, hingga *time series*.
4. Tampilan grafik yang menarik dan fleksibel.
5. Dapat dikembangkan sesuai keperluan dan kebutuhan dan sifatnya yang terbuka, setiap orang dapat menambahkan fitur-fitur tambahan dalam bentuk paket ke dalam *software* R.
6. R bersifat multiplatform, yakni dapat diinstall dan digunakan baik pada *system* operasi *Windows*, *UNIX*/*LINUX* maupun pada *Macintosh*. Untuk dua *system* operasi disebeutkan terakhir diperlukan sedikit penyesuaian.

R (juga dikenal sebagai GNU S) adalah bahasa pemrograman dan perangkat lunak untuk analisis statistika dan grafik (John, 2005). R dibuat oleh Ross Ihaka dan Robert Gentleman di Universitas Auckland, Selandia Baru, dan kini dikembangkan oleh R Development Core Team, di mana Chambers merupakan anggotanya. R dinamakan sebagian setelah nama dua pembuatnya (Robert Gentleman dan Ross Ihaka), dan sebagian sebagian dari permainan nama dari S. (Hornik, 2020)

Bahasa R kini menjadi standar *de facto* di antara statistikawan untuk pengembangan perangkat lunak statistika, serta digunakan secara luas untuk pengembangan perangkat lunak statistika dan analisis data (Vance, 2009). R merupakan bagian dari proyek GNU. Kode sumbernya tersedia secara bebas di bawah Lisensi Publik Umum GNU, dan versi biner prekompilasinya tersedia

untuk berbagai sistem operasi. R menggunakan antarmuka baris perintah, meski beberapa antarmuka pengguna grafik juga tersedia. (Foundation)

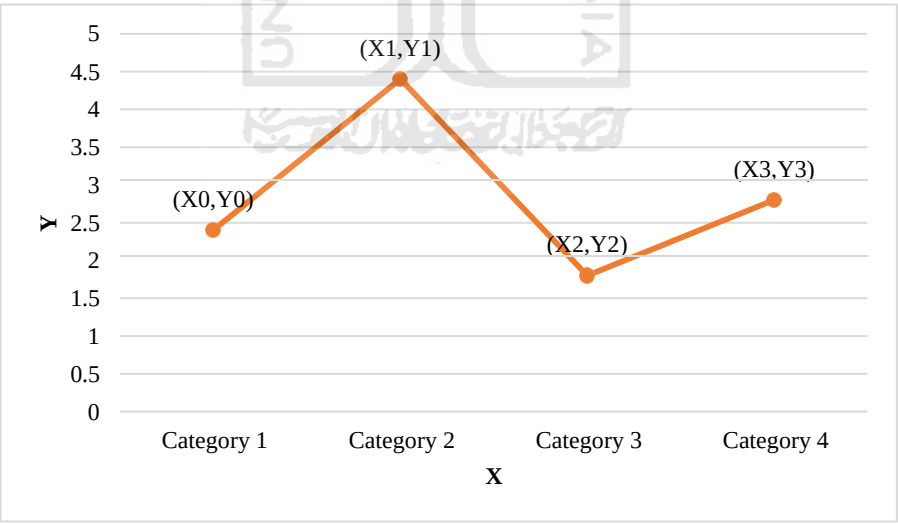
3.5 Interpolasi Linier

Interpolasi adalah pencarian sebuah titik di antara dua sumbu (min) dan (max) yang dalam sebuah data linier disebut interpolasi linier. Interpolasi disebut juga penyisipan atau penambahan data yang tidak tersedia/hilang (Heri Mulyono). Interpolasi adalah metode yang sangat berguna saat kita tidak memperoleh nilai data pada periode tertentu karena data tidak tersedia dengan berbagai penyebab. Cara menentukan nilai fungsi f dititik X^* ε $[X_0, X_n]$ dengan menggunakan informasi dari seluruh atau sebagian titik-titik yang diketahui $(X_0, X_1, \ldots, X_n)$. (Mohamad Sidiq)

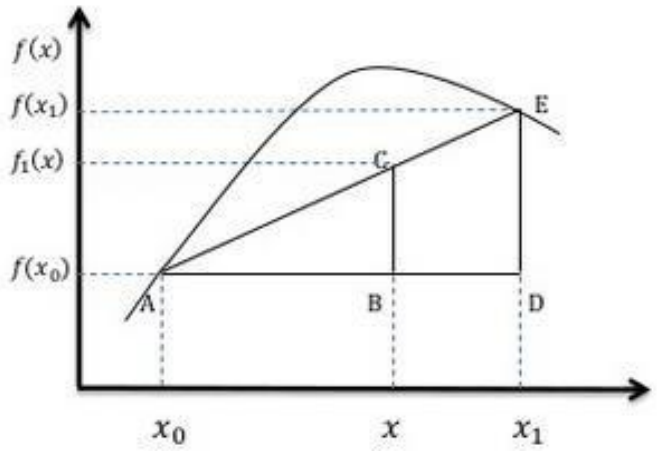
Tabel 3.2 Penentuan Fungsi pada Interpolasi

X	X_0	X_1	X_2	…	X_n
$F(x)$	$F(X_0)$	$F(X_1)$	$F(X_2)$	…	$F(X_n)$

Bentuk paling sederhana dari interpolasi adalah menghubungkan dua titik data dengan garis lurus.



Gambar 3.1 Grafik Sederhana dari Interpolasi



Gambar 3.2 Grafik Sederhana dari Interpolasi

(source : https://www.slideshare.net/oktiagung/metode-interpolasi-linier)

Metode ini disebut dengan interpolasi linier yang dapat dijelaskan dengan rumus sebagai berikut:

— — (3.1)

atau

( ) ( ) ( ) ( ) (3.2)

( ) ( )

sehingga

( ) ( ) ( ) ( ) (3.3)

( )

Data yang hilang adalah masalah yang sangat umum untuk semua jenis data. (Moritz S. , 2017). Dalam R, terdapat sebuah *packages* yang dapat mengatasi data *missing*. *Packages* imputeTS merupakan kumpulan algoritma dan alat untuk imputasi deret waktu univariat. *Packages* ini terdapat beberapa implementasi algoritma imputasi yang berbeda. Selain algoritme imputasi, *packages* ini juga berfungsi merencanakan dan mencetak data yang hilang. *Packages* ini mudah digunakan (Moritz, 2018):

a. Untuk mengisi data yang hilang dalam data deret waktu
   na.interpolation(x)
b. Untuk memplotkan data deret waktu yang hilang
   plotNA.distribution(x)
c. Untuk mengetahui hasil data deret waktu
   statsNA(x)

3.6 Data Mining

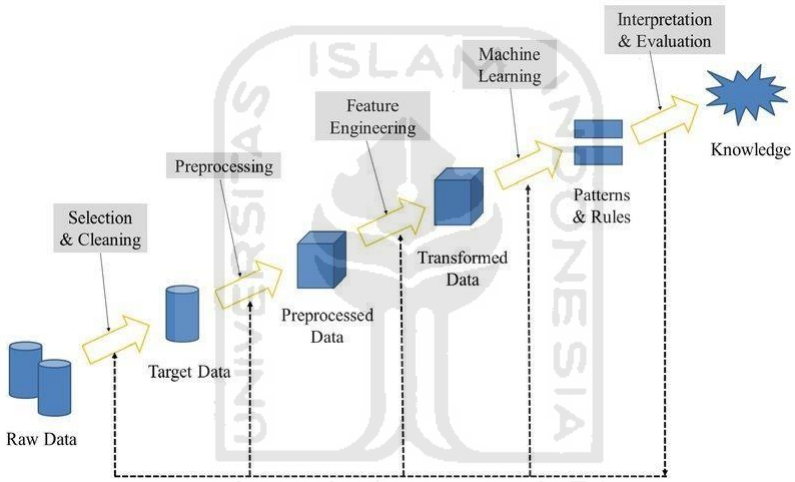
Data Mining merupakan proses pengumpulan dan menggali data untuk menemukan pola yang menarik dari data dalam jumlah besar. Metode ini merupakan gabungan dari *machine learning*, statistik dan sistem basis data. (Christopher, 2010). *Data mining* dapat diterapkan pada berbagai bidang yang mempunyai sejumlah data. *Data mining* bertujuan untuk memperbaiki teknik tradisional, sehingga bisa menangani (Fajrin, 2018):

a. Jumlah data yang sangat besar

b. Dimensi data yang tinggi

c. Data yang heterogen dan berbeda sifat

Dibawah ini merupakan tahapan dari data mining (source : www.google.com)



Gambar 3.3 Tahapan Data Mining

3.7 Ensemble Subset

Misalkan $L = (Xi, Yi)$, dengan $i = 1,\ldots,n$ merupakan himpunan data training yang terdiri dari n pengamatan, dengan $Xi = (X_{i1}, X_{i2}, \ldots, X_{id})$ adalah sebuah vektor variabel berdimensi *d* dan *y* adalah sebuah vektor dari kelas label. Dimana $Y_i \in \{1,\ldots,J\}$, J adalah jumlah kelas total. Dalam kasus penelitian ini, $Y_i \in \{0,1,2,3,4\}$. Berikut ini merupakan tabel ilustrasi dari metode Ensemble Subset:

	Y_i	X_{i1}	X_{i2}	…	X_{id}
i=1	Y_1	X_{11}	X_{12}	…	X_{1d}
i=2	Y_2	X_{21}	X_{22}	…	X_{2d}
i=3	Y_3	X_{31}	X_{32}	…	X_{3d}

i=4	Y_4	X_{41}	X_{42}	…	X_{4d}
i=5	Y_5	X_{51}	X_{52}	…	X_{5d}
…	…	…	…	…	…
n	Y_n	X_{n1}	X_{n2}	…	X_{nd}

Berdasarkan data set himpunan data L, sebuah *classifier* memprediksikan kelas label untuk observasi dengan fitur pada vektor (*X'*). Kemudian, bagi data training *L* menjadi 2 data bagian yaitu L_t dan L_v. Bagian yang pertama akan kita gunakan untuk konstruksi *classifier* dan bagian yang lain untuk validasi. Untuk memudahkan, himpunan yang digunakan untuk mengkonstruksi model pada L_t didefinisikan sebagai L^*. Misalkan d input fitur di L^* didefinisikan dengan $P = P_1, \ldots, P_d$. Untuk ukuran subset yang diberikan, misalkan *l* dengan $l < d$, sebuah subhimpunan dari fitur P^l di ambil dari *P*. Berdasarkan fitur yang dipilih secara acak, sampel *bootstrap* di ambil dari L^*. Pembelajaran *bootstrap* baru pada himpunan $L^{*(l)}$ terdiri dari vektor fitur berdimensi *l*. Proses ini diulang sampai kita mendapatkan *n training* set, $L^{*(1l)}, \ldots, L^{*(m\,l)}$, masing-masing dari ($n$×1+1) dimensi. Dasar *kNN classifier* dibangun dari himpunan data *training* dan sebuah himpunan dari *m classifier* yang dihasilkan.

Sementara, dengan membuat sebuah sampel random dengan ukuran yang sama (*n*) dari himpunan training, di ambil 1/3 dari sampel tersebut. Observasi tersebut kita sebut sebagai *OOB* (*out-of-bag*) klasifikasi dan dapat digunakan untuk mengestimasi kesalahan klasifikasi. Dalam penelitian ini, digunakan sampel *OOB* yang didefinisikan sebagai data *testing* yang digunakan sebagai penguji dari hasil klasifikasi. *m classifier* kemudian diranking sesuai dengan akurasi klasifikasi masing-masing. Klasifikasi tersebut kemudian digunakan untuk melakukan asesmen pada himpunan data L_v. (Gul, 2016)

Pembentukan *ensemble* subset dari pengklasifikasi *kNN* dapat diringkas sebagai berikut :

1. Buat sampel acak ukuran $l < d$, tanpa pengembalian, vektor fitur *P* dari L^*, tentukan vektor fitur P^l
2. Berdasarkan subset acak yang dipilih, buat sampel acak berukuran *n*, $L^{*(l)}$ dari L^*

3. Susunlah *kNN classifier* pada $L^{*(l)}$
4. Hitung keakuratan klasifikasi untuk sampel *OOB*, menggunakan variabel yang sama seperti yang digunakan untuk konstruksinya.
5. Ulangi langkah ke 1-4 hingga *m* kali dan beri peringkat *m* pengklasifikasi berdasarkan keakuratannya.
6. Pilih *h* pengklasifikasi pertama dengan akurasi tertinggi

3.8 K-Nearest Neighbor

Dalam pola rekognisi, algoritma kNN adalah metode non parametrik yang digagas oleh Thomas Cover yang digunakan untuk klasifikasi dan regresi. Dalam dua hal tersebut, input memiliki data *training* k terdekat dalam ruang fitur. Outputnya bergantung pada apakah kNN digunakan untuk klasifikasi atau regresi. Dalam klasifikasi kNN, output adalah sebuah keanggotaan kelas. Sebuah objek diklasifikasikan berdasarkan perbedaan pilihan tetangganya (k adalah bilangan bulat positif, biasanya kecil). Jika k=1, maka objek ditempatkan sebagai kNN tunggal. (Altman, 1992)

kNN adalah suatu tipe pembelajaran berbasis contoh dimana fungsi hanya diberi pendekatan terbatas dan semua perhitungan dilakukan setelah evaluasi fungsi. Oleh karena itu, algoritma ini mendasarkan pada jarak bagi klasifikasi, menormalkan data *training* dan dapat meningkatkan keakuratannya secara dramatis. (Altman, 1992). Algoritma *k-Nearest Neighbor* (kNN) merupakan sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut. kNN termasuk algoritma *supervised learning* dimana hasil dari *query instance* yang baru diklasifikan berdasarkan mayoritas dari kategori pada kNN. Kelas yang paling banyak muncul itu yang akan menjadi kelas hasil klasifikasi. Tujuan dari algoritma ini adalah mengklasifikasikan objek baru berdasarkan atribut dan sampel-sampel dari data *training*. (Banjarsari, 2015).

Prinsip kerja kNN adalah mencari jarak terdekat antara data yang akan dievaluasi dengan k tetangga terdekatnya dalam data *training*, dengan k merupakan banyaknya tetangga terdekat. Data *training* diproyeksikan ke dalam ruang berdimensi banyak, dimana masing-masing dimensi menjelaskan fitur dari

data. Ruang ini dibagi menjadi bagian-bagian berdasarkan klasifikasi data *training*. (Ramadhani, 2019)

Langkah yang digunakan dalam metode *k-Nearest Neighbor* yaitu (Budiman, 2015) :

1. Tentukan parameter k (jumlah tetangga paling dekat).
2. Hitung jarak *Euclid* masing-masing objek terhadap data sampel yang diberikan.
3. Urutkan seluruh jarak berdasarkan jarak minimum dan tetapkan sesuai dengan nilai k.
4. Kumpulkan kategori Y (klasifikasi *nearest neighbor*).
5. Gunakan kelas dengan jumlah terbanyak untuk menentukan nilai *query instance* yang telah dihitung.

Untuk menentukan jauh atau dekatnya jarak antar titik pada kelas k dapat dihitung menggunakan rumus jarak *Euclidean*. Jarak *Euclidean* adalah formula untuk mencari jarak antara 2 titik dalam ruang dua dimensi. (Ramadhani, 2019)

$$( ) \sqrt{\sum}( ) \qquad (3.4)$$

dimana:

d (*x*, y) : jarak *Euclidean*

n : banyaknya data

i : 1,2,3,…, *n*

kNN memilih *k* data dari data *training* yang dekat dengan data *testing* dalam memprediksi peubah *output*. Nilai *output* dari *k* data *training* yang terpilih sebagai tetangga terdekat digunakan untuk memprediksi nilai *output* dari data *testing* yang tidak diketahui. Pada penelitian ini prediksi menggunakan modifikasi metode kNN yaitu dengan menambahkan faktor koreksi *trend* dan perubahan waktu yang di gambarkan pada persamaan (3.5) (Sorjamaa A, 2005).

$$\frac{\sum}{} \qquad (3.5)$$

dimana:

y_i : data hasil kNN

w_j : pembobot untuk tetangga ke-*j*

y_j : peubah *output* ke-*j*

b : *slope*

D : rata-rata selisih antara nomor urut data *testing* dengan data *training* yang terpilih menjadi *k* tetangga terdekat

k : banyaknya tetangga terdekat ke-*n*

j : 1,2,3,…,*k*

Pembobot ini dapat disesuaikan berdasarkan data yang diamati, yaitu *wj* = *q n* dengan *q* = urutan waktu dan *n* = banyaknya amatan pada data *training*.

3.9 Confussion Matrix

Confussion matrix melakukan pengujian untuk memperkirakan obyek yang benar dan salah (F, 2011). Urutan pengujian ditabulasikan dalam *confussion matrix* dimana kelas yang diprediksi ditampilkan di bagian atas matriks dan kelas yang diamati di bagian kiri. Setiap sel berisi angka yang menunjukkan berapa banyak kasus yang sebenarnya dari kelas yang diamati untuk diprediksi. (Menarianti, 2015)

Tabel 3.1 *Confussion* Matrix

Classification		Predicted Class	
		Class = YES	Class = NO
Observed Class	Class = YES	TP	FN
	Class = NO	FP	TN

Keterangan :

TP = *True Positive*

FP = *False Positive*

FN = *False Negative*

TN = *True Negative*

Berikut macam-macam dari *confussion matrix*:

a. Recall

Recall merupakan metode pengujian yang membandingkan jumlah informasi relevan yang didapatkan sistem dengan jumlah seluruh informasi relevan yang ada

dalam koleksi informasi (baik yang terambil atau tidak terambil oleh sistem). Persamaan *recall* ditunjukkan pada persamaan berikut.

$$Recall = \frac{\quad}{\quad}$$

b. Precision

Presisi merupakan metode pengujian dengan melakukan perbandingan jumlah informasi relevan yang didapatkan sistem dengan jumlah seluruh informasi yang terambil oleh sistem baik yang relevan maupun tidak. Persamaan presisi ditunjukkan pada persamaan berikut.

$$Precision = \frac{\quad}{\quad}$$

c. Accuracy

Akurasi merupakan metode pengujian berdasarkan tingkat kedekatan antara nilai prediksi dengan nilai aktual. Persamaan akurasi ditunjukkan pada persamaan berikut.

$$Accuracy = \frac{\quad}{\quad}$$

d. AUC dan ROC

ROC (*Receiver Operating Characteristic*) *curve* banyak digunakan dalam penelitian data mining dalam menilai hasil prediksi (F, 2011). Secara teknis ROC *curve* dibagi dalam dua dimensi, dimana tingkat TP diletakkan pada sumbu Y dan tingkat FP diletakkan pada sumbu X. Tetapi untuk merepresentasikan grafis yang menentukan klasifikasi mana yang lebih baik, digunakan metode yang menghitung luas daerah dibawah ROC yang disebut AUC (Area Under the Curve) yang diartikan sebagai probabilitas (Witten, 2011). AUC mengukur kinerja diskriminatif dengan memperkirakan probabilitas *output* dari sampel yang dipilih secara acak dari populasi positif atau negatif. Semakin besar AUC, semakin kuat klasifikasi yang digunakan (Yu, 2007). Berikut ini tingkat keakuratan klasifikasi dengan menggunakan AUC: (Suwarno, 2016)

Tabel 3.2 Interpretasi Nilai AUC

Nilai AUC	**Interpretasi**
0.90 – 1.00	*Excellent classification*

Nilai AUC	Interpretasi
0.80 – 0.90	*Good classification*
0.70 – 0.80	*Fair classification*
0.60 – 0.70	*Poor classification*
0.50 – 0.60	*Failure*

e. Kappa Statistic

Statistik kappa sering digunakan untuk menguji reliabilitas antar penilai. Pengukuran sejauh mana pengumpul data (penilai) menetapkan skor yang sama untuk variabel yang sama disebut reliabilitas antar penilai. Seperti kebanyakan statistik korelasi, kappa dapat berkisar dari -1 hingga +1. Walaupun kappa adalah salah satu statistik yang paling sering digunakan untuk menguji reliabilitas antar penilai, kappa juga memiliki keterbatasan.
(https://www.ncbi.nlm.nih.gov/pubmed/23092060)

Berikut merupakan tabel nilai interpretasi kappa: (Anthony, 2005)

Tabel 3.3 Interpretasi Nilai Kappa

Nilai Kappa	Interpretasi
< 0	*Less than chance agreement*
0.01 – 0.20	*Slight agreement*
0.21 – 0.40	*Fair agreement*
0.41 – 0.60	*Moderate agreement*
0.61 – 0.80	*Substantial agreement*
0.81 – 0.99	*Almost perfect agreement*

BAB IV
METODOLOGI PENELITIAN

4.1 Populasi dan Sampel

Populasi dalam penelitian ini adalah curah hujan harian di seluruh Indonesia tahun 2014 sampai 2018, sedangkan sampel yang digunakan dalam penelitian ini adalah curah hujan harian di Kota Bogor pada tanggal 01 Januari 2014 sampai 31 Desember 2018 pada Stasiun Klimatologi Bogor.

4.2 Sumber Data

Data yang digunakan dalam penelitian ini adalah data sekunder yang bersumber dari *website* resmi BMKG Indonesia yaitu (BMKG). (Tanggal akses : 5 Januari 2019).

4.3 Variabel Penelitian

Penelitian ini menggunakan dua macam variabel yaitu variabel independen dan variabel dependen sebagai berikut:

Tabel 4.1 Variabel Penelitian

Jenis Variabel	Nama Variabel	Definisi operasional
Variabel Independen (X)	Rata-rata uhu C	Rata-rata suhu setiap hari dalam satuan C yang diperoleh dari *website* bmkg.go.id
	Rata-rata Kelembaban (%)	Rata-rata kelembaban setiap hari dalam persentase yang diperoleh dari *website* bmkg.go.id
	Lama Penyinaran (jam)	Lama penyinaran matahari setiap hari dalam satuan jam yang diperoleh dari *website* bmkg.go.id
	Arah Angin Saat Kecepatan Maksimum (deg)	Arah angin saat kecepatan maksimum setiap hari dalam satuan deg yang diperoleh dari *website* bmkg.go.id
Variabel Dependen (Y)	Curah Hujan (mm) dengan kategori :	0 = sangat ringan 1 = ringan 2 = sedang 3 = lebat 4 = sangat lebat

4.3 Metode Pengumpulan Data

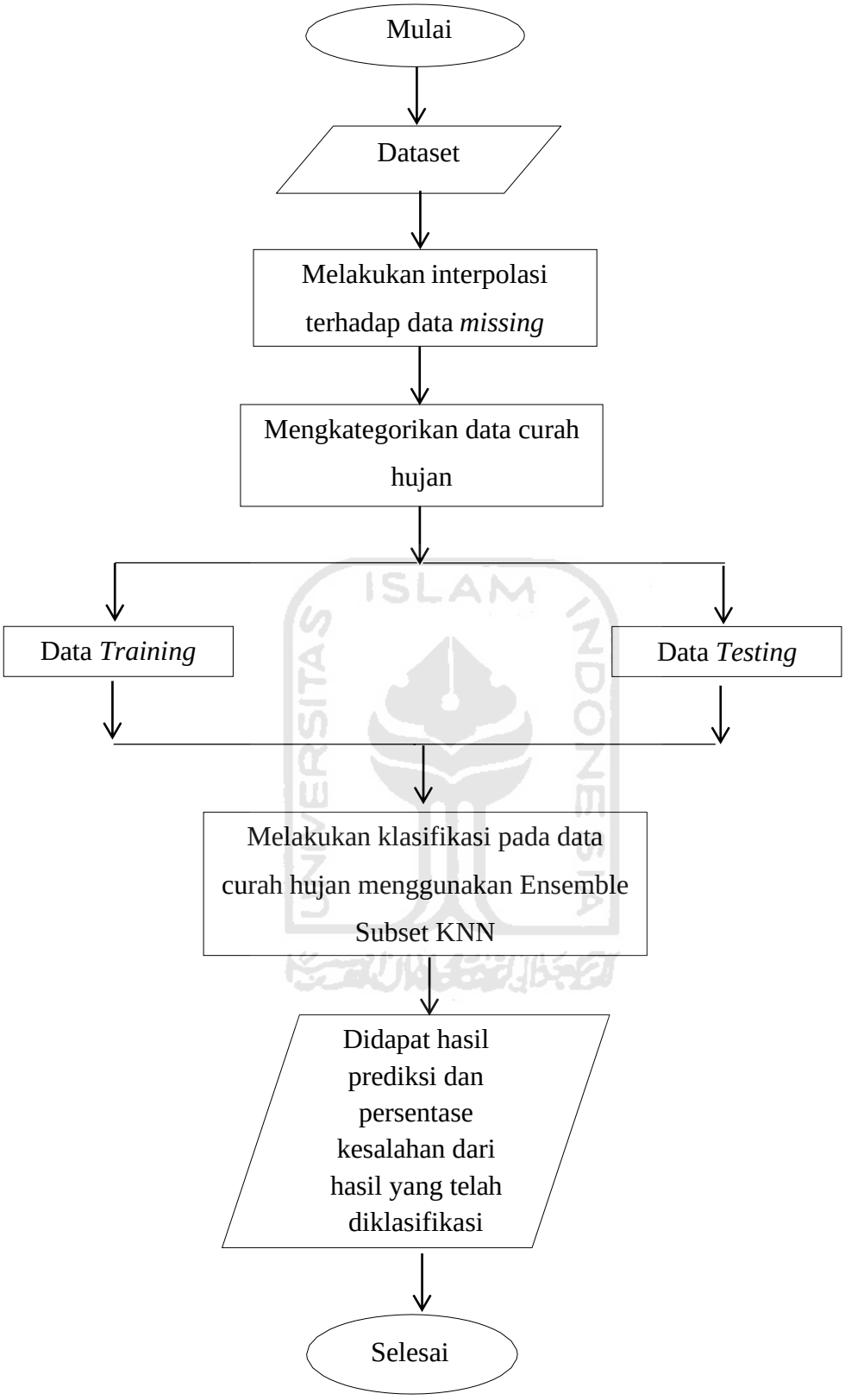
Pada penelitian ini, data yang digunakan merupakan data sekunder yang diperoleh dari *website* resmi Badan Meteorologi dan Geofisika (BMKG) Indonesia. (Tanggal akses : 5 Januari 2019).

4.4 Metode Analisis Data

Penelitian ini menggunakan metode *Ensemble* Subset kNN yang kemudian akan diperoleh hasil klasifikasi curah hujan, persentase kesalahan serta nilai ketepatan klasifikasi.

4.5 Flowchart

Adapun tahapan penelitian yang dilakukan oleh peneliti seperti pada *flowchart* berikut:



Gambar 4.1 *Flowchart* Penelitian

BAB V
PEMBAHASAN

Pada Bab V akan dibahas terkait hasil dari penelitian yang meliputi preprosesing data yang terdiri dari penanganan *missing value*, pengklasifikasian jenis curah hujan dan dilanjutkan dengan Statistika Deskriptif. Kemudian dilanjutkan dengan pemaparan hasil implementasi ESKnn, hasil ketepatan klasifikasi dan selesai. Berikut merupakan hasil dari pembahasan penelitian.

5.1 Statistika Deskriptif

Sebelum melakukan visualisasi data, berdasarkan sumber data yang digunakan masih terdapat *missing value* sehingga perlu dilakukan tahapan awal yaitu penanganan *missing value* agar didapatkan hasil yang optimal untuk peramalannya. Penanganan *missing value* ini menggunakan *packages* imputeTS serta perintah na.interpolation pada aplikasi RStudio.

Berikut merupakan *input* dari data awal yang masih terdapat *missing value*.

Tabel 5.1 Data yang masih terdapat Missing Value

Pengamatan ke	**Y (Curah Hujan)**	**X_1 (Suhu rata-rata)**	**X_2 (Kelembaban rata-rata)**	**X_3 (Lama Penyinaran)**	**X_4 (Arah Angin)**
1	47.7	24.4	91	0	315
2	3.8	24.1	94	5.7	NA
3	NA	NA	NA	5.3	315
4	43.2	24.7	91	NA	225
5	21	24.7	88	5.4	270

Dapat dilihat pada tabel 5.1, peneliti mengambil masing-masing satu contoh *missing value* yang terdapat di seluruh variabel. Pada data tersebut terdapat missing yang ditandai dengan tulisan NA. Untuk menangani data *missing* tersebut agar dapat dianalisis, maka dilakukan interpolasi dengan menggunakan perintah na.Interpolation beserta *package* imputeTS sehingga didapatkanlah hasil untuk menggantikan nilai NA tersebut dengan angka real yang nantinya dapat dianalisis. Berikut merupakan hasil dari penanganannya.

Tabel 5.2 Data yang bersih dari Missing Value

Pengamatan ke	Y (Curah Hujan)	X_1 (Suhu rata-rata)	X_2 (Kelembaban rata-rata)	X_3 (Lama Penyinaran)	X_4 (Arah Angin)
1	47.7	24.4	91	0	315
2	3.8	24.1	94	5.7	315
3	23.5	24.4	92.5	5.3	315
4	43.2	24.7	91	5.35	225
5	21	24.7	88	5.4	270

Dapat dilihat pada tabel 5.2, terlihat beberapa kotak berwarna kuning yang artinya, *missing value* telah diatasi. Berikut ini merupakan salah satu contoh penjabaran dari perhitungan interpolasi.

Data ke-4 ~> 43,2

Data ke-2 ~> 3,8

———+

47,0 ÷ 2 = 23,5 mm

Perhitungan diatas mengambil contoh pada salah satu data dalam variabel curah hujan, seperti dapat dilihat pada tabel 5.2. Kotak pada tabel yang diberi tanda warna kuning merupakan *output* dari aplikasi R dan perhitungan diatas merupakan pembuktian serta penjabaran secara manual proses interpolasi yang dilakukan pada data curah hujan.

Untuk data selengkapnya dapat dilihat pada lampiran 2

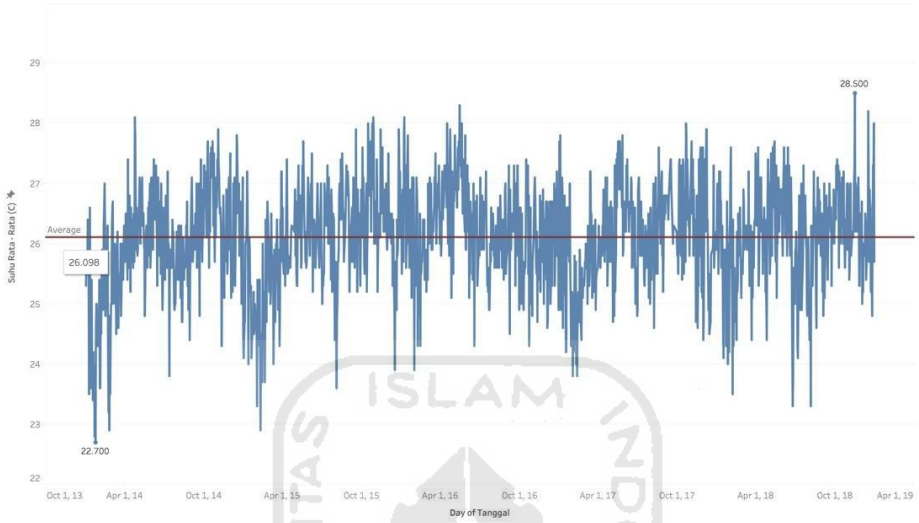
Proses selanjutnya yaitu pengkategorian data curah hujan berdasarkan penelitian yang dilakukan oleh (Fauziah, 2008) terdapat di landasan teori Tabel 3.1 didapatkan hasil pengkategorian sebagai berikut.

Tabel 5.3 Pengkategorian Data Curah Hujan

Curah Hujan	Label	Status
0	0	Sangat Ringan
9.2	1	Ringan
36.8	2	Sedang
84.6	3	Lebat
102.2	4	Sangat Lebat

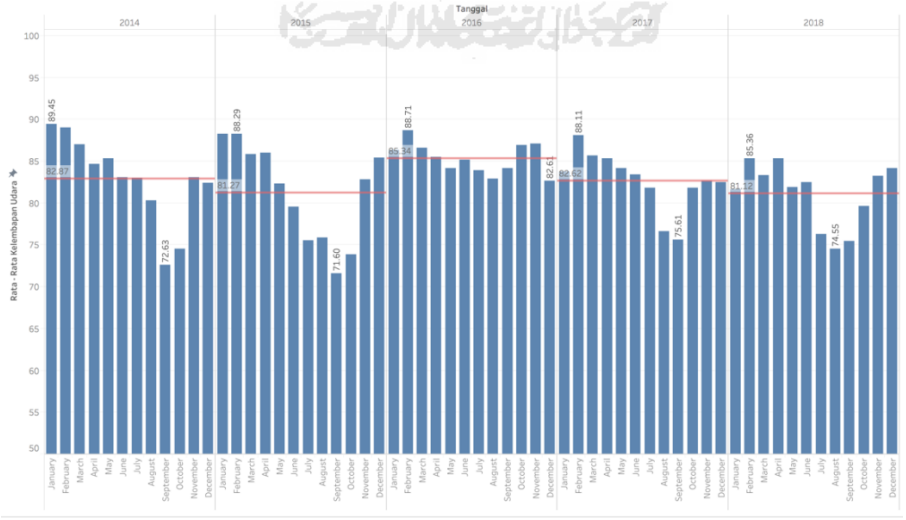
Untuk data selengkapnya dapat dilihat pada lampiran 3.

Proses selanjutnya adalah visualisasi data untuk masing-masing variabel yang akan ditampilkan dalam bentuk grafik. Berikut visualisasi dari masing-masing variabelnya.



Gambar 5.1 Grafik Waktu dan Suhu Rata-Rata

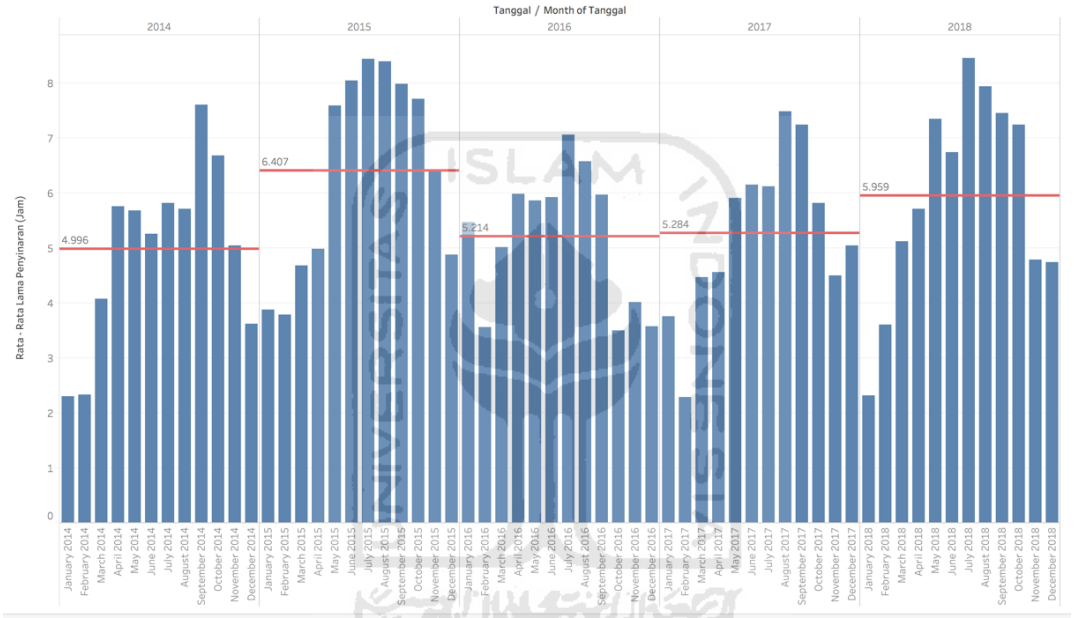
Berdasarkan gambar 5.1 dapat dilihat bahwa suhu terendah terdapat pada tahun 2014 dengan angka 22.7◦ dan suhu tertinggi terdapat pada tahun 2018 dengan angka 2 .5◦. Rata-rata suhu keseluruhan pada tahun 2014 hingga 2018 yaitu 26◦.



Gambar 5.2 Grafik Waktu dan Kelembaban Rata-Rata

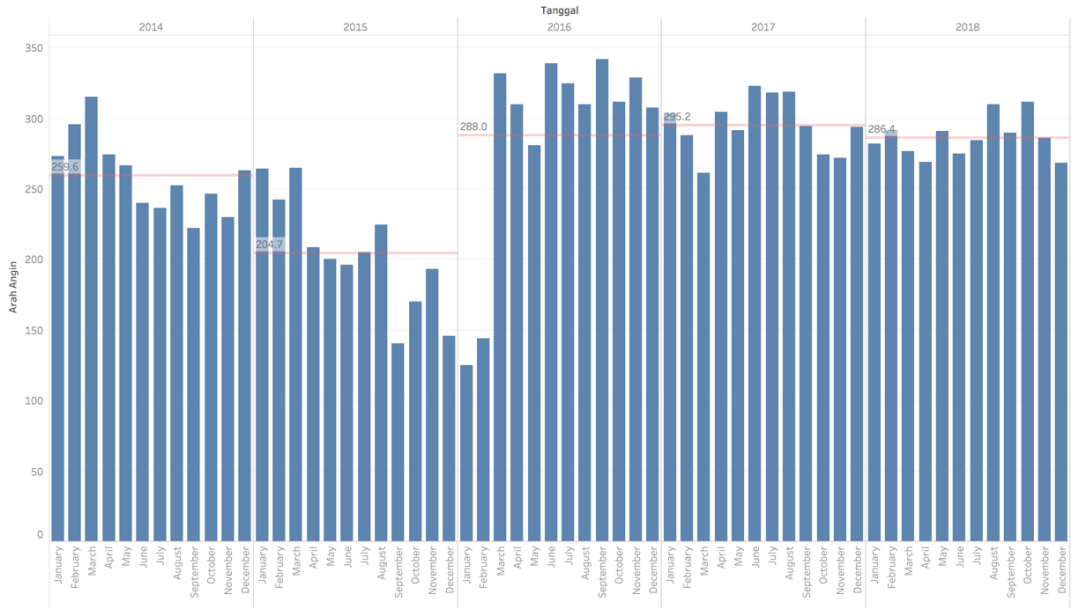
Berdasarkan gambar 5.2 dapat dilihat bahwa pada tahun 2014 kelembaban tertinggi di bulan Januari dengan angka 89.45% dan terendah di bulan September

dengan angka 72.63% serta rata-ratanya 82.87%. Pada tahun 2015 kelembaban tertinggi di bulan Februari dengan angka 88.29% dan terendah di bulan September dengan angka 71.6% serta rata-ratanya 81.27%. Pada tahun 2016 kelembaban tertinggi di bulan Februari dengan angka 88.71% dan terendah di bulan Desember dengan angka 82.61% serta rata-ratanya 85.34%. Pada tahun 2017 kelembaban tertinggi di bulan Februari dengan angka 88.11% dan terendah di bulan September dengan angka 75.61% serta rata-ratanya 82.62%. Pada tahun 2018 kelembaban tertinggi di bulan Februari dengan angka 85.36% dan terendah di bulan Agustus dengan angka 74.55% serta rata-ratanya 81.12%.



Gambar 5.3 Grafik Waktu dan Lama Penyinaran

Berdasarkan gambar 5.3 dapat dilihat bahwa pada tahun 2014 rata-rata lama penyinaran selama 4 jam. Tahun 2015 rata-rata lama penyinaran selama 6 jam. Tahun 2016, 2017 dan 2018 rata-rata lama penyinaran selama 5 jam.



Gambar 5.4 Grafik Waktu dan Arah Angin

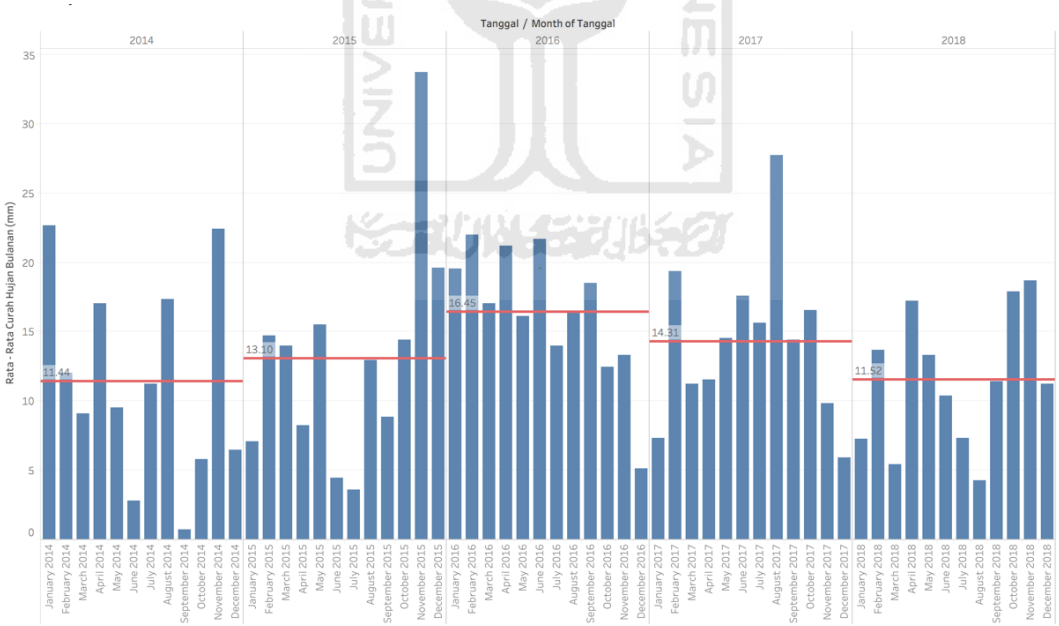
Berdasarkan gambar 5.4 dapat dilihat bahwa pada tahun 2014 rata-rata arah angin sebesar 259.6 deg. Tahun 2015 rata-rata arah angin sebesar 204.7 deg. Tahun 2016 rata-rata arah angin sebesar 288 deg. Tahun 2017 rata-rata arah angin sebesar 295.2 deg. Tahun 2018 rata-rata arah angin sebesar 286.4 deg.

Dalam hal ini, pada tahun 2016 terdapat Dua Siklon Tropis (TC) terbentuk di dekat wilayah Indonesia, yaitu TC YVETTE dan TC NOCK-TEN. TC YVETTE tumbuh di Samudera Hindia Sebelah Selatan Jawa, pada posisi 14.1 BT, 114.2 LS, tepatnya pada jarak sekitar 653 km dari Denpasar pada hari Rabu (21 Des 2016, 13.00 WIB). Pusat tekanan rendahnya 990 hPa dan kecepatan angin maksimum di sekitarnya 75 km/jam. Pergerakan TC YVETTE perlahan ke arah timur- tenggara menjauhi wilayah Indonesia. Sedangkan TC NOCK-TEN terbentuk di Samudera Pasifik Barat sebelah utara Papua, tepatnya di 8.1 LU, 139.6 BT pada hari Kamis (22 Desember 2016, 04.00 WIB). Pusat tekanan rendah TC NOCK-TEN mencapai 1000 hPa dengan angin maksimum di sekitarnya mencapai 93 km/jam dan pergerakannya ke arah Barat Laut dengan kecepatan 20 km/jam. Dalam waktu dua hari ke depan kedua TC tersebut masih dapat terbentuk.

Terbentuknya dua TC tersebut secara tidak langsung dapat menyebabkan kondisi cuaca signifikan di beberapa wilayah di Indonesia. Pola siklonik yang terbentuk dari dua TC tersebut menyebabkan pola angin di wilayah Indonesia

cukup signifikan memberikan dampak pada pembentukan awan hujan dengan potensi hujan sedang hingga lebat di beberapa wilayah dan angin kencang terutama di wilayah perairan yang menyebabkan timbulnya gelombang tinggi.

Dalam periode tiga hari kedepan perlu di waspadai potensi angin kencang di wilayah Laut Jawa, Laut Bali, sebagian wilayah Jawa Barat hingga Jawa Timur, Pesisir selatan Jawa Tengah, Jawa Timur, Bali, NTB, NTT. Potensi hujan sedang hingga lebat di wilayah Jawa Timur bagian Timur dan Selatan, Bali, NTB, NTT, Papua Barat, Papua bagian Utara, dan Biak. Potensi gelombang laut dengan ketinggian 2.0 - 2,5 meter dapat terjadi di samudera Pasifik Barat sebelah utara Biak, Samudera Pasifik sebelah utara Jayapura, perairan Biak dan Perairan Jayapura- Sarmi. Gelombang laut dengan ketinggian 2.5 - 4.0 meter dapat terjadi di wilayah Laut Jawa bagian tengah dan timur, Samudera Hindia selatan Jawa Tengah hingga Jawa Timur, Perairan selatan Jawa Tengah hingga NTB, selat Bali bagian selatan, Laut Sumbawa, Laut Flores bagian barat. Sedangkan potensi gelombang laut dengan ketinggian lebih dari 4.0 meter dapat terjadi di wilayah Samudera Hindia selatan Bali hingga NTT. (BMKG, 2016)



Gambar 5.5 Grafik Waktu dan Curah Hujan

Berdasarkan gambar 5.5 dapat dilihat bahwa curah hujan pada tahun 2014 hingga 2018 tidak menentu. Seperti di bulan September 2014 curah hujannya sangat sedikit. Hal ini disebabkan pada bulan tersebut menurut Kepala Badan Penanggulangan Bencana Daerah (BPBD) Bogor mengatakan bahwa wilayah

Bogor darurat kekeringan sejak awal hingga akhir bulan September 2014. Selain itu di bulan November 2015 curah hujannya sangat tinggi. Hal ini disebabkan pada bulan tersebut menurut Kepala Stasiun Klimatologi Dramaga Bogor mengatakan terjadi cuaca ekstrem hingga bulan Desember 2015 akibat adanya peralihan dari musim kemarau hingga musim penghujan. Namun secara keseluruhan, rata-rata curah hujan dari tahun 2014-2018 berada pada kisaran 11-16 mm. Jumlah hari berdasarkan kategori curah hujan didapatkan :

Tabel 5.4 Jumlah Hari Berdasarkan Status Curah Hujan

Tahun	Bulan	Status Curah Hujan (hari)				
		Sangat Ringan	Ringan	Sedang	Lebat	Sangat Lebat
2014	Jan	10	10	8	1	2
	Feb	15	5	8	0	0
	Mar	18	6	6	1	0
	Apr	16	7	4	2	1
	Mei	17	9	5	0	0
	Jun	25	3	2	0	0
	Jul	21	4	4	1	1
	Ags	20	4	3	2	2
	Sept	27	3	0	0	0
	Okt	25	4	1	1	0
	Nov	15	4	8	2	1
	Des	24	4	2	1	0
2015	Jan	14	11	3	0	0
	Feb	13	8	6	1	0
	Mar	11	11	6	1	0
	Apr	17	5	3	0	0
	Mei	11	8	9	1	0
	Jun	24	3	0	1	0
	Jul	27	0	1	1	0
	Ags	24	1	2	4	0
	Sept	17	11	0	2	0
	Okt	21	1	8	1	0
	Nov	5	7	12	5	1
	Des	7	15	5	4	0
2016	Jan	14	5	10	1	1
	Feb	8	11	6	2	1
	Mar	13	7	8	3	0
	Apr	11	7	8	3	1
	Mei	11	11	7	2	0
	Jun	4	15	8	3	0

Tahun	Bulan	Status Curah Hujan (hari)				
		Sangat Ringan	Ringan	Sedang	Lebat	Sangat Lebat
	Jul	11	14	5	1	0
	Ags	9	13	7	2	0
	Sept	10	11	7	2	0
	Okt	14	10	6	1	0
	Nov	12	12	6	0	0
	Des	18	13	0	0	0
2017	Jan	19	8	4	0	0
	Feb	3	14	10	1	0
	Mar	10	14	5	0	0
	Apr	16	8	6	0	0
	Mei	15	6	10	0	0
	Jun	15	7	5	2	1
	Jul	16	7	5	3	0
	Ags	4	7	19	1	0
	Sept	7	11	3	1	0
	Okt	6	15	5	2	0
	Nov	14	11	5	0	0
	Des	19	5	3	0	0
2018	Jan	18	10	3	0	0
	Feb	13	9	3	3	0
	Mar	20	9	2	0	0
	Apr	9	9	12	0	0
	Mei	18	7	5	0	1
	Jun	17	9	3	1	0
	Jul	12	19	0	0	0
	Ags	23	7	1	0	0
	Sept	19	5	4	2	0
	Okt	18	4	6	2	1
	Nov	8	11	9	2	0
	Des	16	10	2	3	0

		Curah Hujan				
Year of Tanggal	Quarter of Tanggal	Sangat Lebat	Lebat	Sedang	Ringan	Sangat Ringan
2014	Q1	2	2	22	21	43
	Q2	1	2	11	19	58
	Q3	3	3	7	11	68
	Q4	1	4	11	12	64
	Total	7	11	51	63	233
2015	Q1		2	15	30	38
	Q2		2	12	16	52
	Q3		7	3	12	68
	Q4	1	10	25	23	33
	Total	1	21	55	81	191
2016	Q1	2	6	24	23	35
	Q2	1	8	23	33	26
	Q3		5	19	68	
	Q4		1	12	79	
	Total	3	20	78	203	61
2017	Q1		1	19	68	
	Q2	1	2	21	67	
	Q3		5	27	52	
	Q4		2	39	44	
	Total	1	10	106	231	
2018	Q1		3	59	28	
	Q2	1	1	64	25	
	Q3		2	59	31	
	Q4	10	37	20	25	
	Total	11	43	202	109	

Gambar 5.6 Jumlah Hari Berdasarkan Status Curah Hujan

Berdasarkan Gambar 5.6 didapatkan kesimpulan bahwa hari hujan dengan kriteria sangat lebat paling banyak terjadi di tahun 2018.

Proses selanjutnya akan dilakukan analisis dengan menggunakan ESKnn.

5.2 ESKnn

Langkah pertama dalam penyelesaian menggunakan ESKnn adalah dengan membagi data menjadi data *training* dan data *testing*. Pada penelitian ini menggunakan skenario yaitu 70% atau sebanyak 1.255 merupakan data *training* sedangkan 30% atau sebanyak 538 digunakan sebagai data *testing*. Berikut merupakan hasil pembagian data *training* dan data *testing*.

Tabel 5.5 Data *Training*

Data ke	Suhu rata-rata	Kelembaban rata-rata	Lama Penyinaran	Arah Angin	Curah Hujan	Label
1	25.3	89	0	360	10	1
:						
1255	25.6	86	4.4	270	51.2	3

Tabel 5.6 Data *Testing*

Data ke	Suhu rata-rata	Kelembaban rata-rata	Lama Penyinaran	Arah Angin	Curah Hujan	Label
1256	25.5	85	5.4	360	0	1
:						
1792	25.7	82	8.8	230	1.8	4

Berikut merupakan hasil *output* R berupa tabel *confussion matrix* beserta ketepatan klasifikasi dan nilai kesalahan dengan memasukkan nilai *k* = 3, 4, 5 dan 6.

a. Untuk k=3

Dibawah ini merupakan tabel *confussion matrix* yang dihasilkan dari aplikasi R, jika nilai k-nya 3. Selain itu pula dilakukan penjumlahan dari masing-masing baris dan kolom.

Tabel 5.7 *Confussion* Matrix dengan k=3

Aktual	Prediksi					
0	110	15	19	3	0	147
1	26	156	18	3	0	203
2	16	22	103	6	0	147
3	2	3	1	24	1	31
4	2	1	0	2	5	10
Sum	156	197	141	38	6	538

Dapat dilihat pada tabel 5.7 diatas, ketika status curah hujan aktual 0 diprediksi 0 terdapat 110 kelas, aktual 1 diprediksi 1 terdapat 156 kelas, aktual 2 diprediksi 2 terdapat 103 kelas, aktual 3 diprediksi 3 terdapat 24 kelas dan aktual 4 diprediksi 4 terdapat 5 kelas.

Selanjutnya didapatkan hasil ketepatan klasifikasi yang dibuktikan juga dengan pengerjaan manual pada Ms. Excel. Berikut ini merupakan hasilnya.

Tabel 5.8 *Confussion* Matrix dengan k=3

Aktual	Prediksi				
	0	1	2	3	4
0	0.204461	0.027881	0.035316	0.005576	0
1	0.048327	0.289963	0.033457	0.005576	0

Aktual	Prediksi				
	0	1	2	3	4
2	0.02974	0.40892	0.19145	0.11152	0
3	0.003717	0.005576	0.001859	0.04461	0.001859
4	0.003717	0.001859	0	0.003717	0.009294

Dalam tabel 5.8 diatas terdapat 5 kotak yang berwarna hijau, apabila kelima nilai tersebut dijumlahkan, didapatkan hasil ketepatan klasifikasi 0.739777. Untuk mendapatkan nilai kesalahan, maka 1 – 0.739777 = 0.260223 yang berarti kesalahannya sebesar 26%.

b. Untuk k=4

Dibawah ini merupakan tabel *confussion matrix* yang dihasilkan dari aplikasi R, jika nilai k-nya 4. Selain itu pula dilakukan penjumlahan dari masing-masing baris dan kolom.

Tabel 5.9 *Confussion* Matrix dengan k=4

Aktual	Prediksi					
0	115	13	17	2	0	147
1	31	152	15	5	0	203
2	15	23	103	6	0	147
3	2	3	6	19	1	31
4	1	2	0	3	4	10
Sum	164	193	141	35	5	538

Dapat dilihat pada tabel 5.9 diatas, ketika status curah hujan aktual 0 diprediksi 0 terdapat 115 kelas, aktual 1 diprediksi 1 terdapat 152 kelas, aktual 2 diprediksi 2 terdapat 103 kelas, aktual 3 diprediksi 3 terdapat 19 kelas dan aktual 4 diprediksi 4 terdapat 4 kelas.

Selanjutnya didapatkan hasil ketepatan klasifikasi yang dibuktikan juga dengan pengerjaan manual pada Ms. Excel. Berikut ini merupakan hasilnya.

Tabel 5.10 *Confussion* Matrix k=4

Aktual	Prediksi				
	0	1	2	3	4
0	0.213755	0.024164	0.031599	0.003717	0
1	0.057621	0.282528	0.027881	0.009294	0
2	0.027881	0.042751	0.19145	0.011152	0
3	0.003717	0.005576	0.11152	0.035316	0.001859
4	0.001859	0.003717	0	0.005576	0.007435

Dalam tabel 5.10 diatas terdapat 5 kotak yang berwarna hijau, apabila kelima nilai tersebut dijumlahkan, didapatkan hasil ketepatan klasifikasi 0.730483. Untuk mendapatkan nilai kesalahan, maka 1 – 0.730483 = 0.2695167 yang berarti kesalahannya sebesar 26%.

c. Untuk k=5

Dibawah ini merupakan tabel *confussion matrix* yang dihasilkan dari aplikasi R, jika nilai k-nya 5. Selain itu pula dilakukan penjumlahan dari masing-masing baris dan kolom.

Tabel 5.11 *Confussion* Matrix dengan k=5

Aktual	Prediksi					
0	112	17	16	2	0	147
1	31	157	13	2	0	203
2	15	26	101	5	0	147
3	4	2	6	19	0	31
4	2	1	0	3	4	10
Sum	164	203	136	31	4	538

Dapat dilihat pada tabel 5.11 diatas, ketika status curah hujan aktual 0 diprediksi 0 terdapat 112 kelas, aktual 1 diprediksi 1 terdapat 157 kelas, aktual 2 diprediksi 2 terdapat 101 kelas, aktual 3 diprediksi 3 terdapat 19 kelas dan aktual 4 diprediksi 4 terdapat 4 kelas.

Selanjutnya didapatkan hasil ketepatan klasifikasi yang dibuktikan juga dengan pengerjaan manual pada Ms. Excel. Berikut ini merupakan hasilnya.

Tabel 5.12 *Confussion* Matrix k=5

Aktual	Prediksi				
	0	1	2	3	4
0	0.208178	0.031599	0.02974	0.003717	0
1	0.057621	0.291822	0.024164	0.003717	0
2	0.027881	0.048327	0.187732	0.009294	0
3	0.007435	0.003717	0.011152	0.035316	0
4	0.003717	0.001859	0	0.005576	0.007435

Dalam tabel 5.12 diatas terdapat 5 kotak yang berwarna hijau, apabila kelima nilai tersebut dijumlahkan, didapatkan hasil ketepatan klasifikasi 0.730483. Untuk mendapatkan nilai kesalahan, maka 1 – 0.730483 = 0.2695167 yang berarti kesalahannya sebesar 26%.

d. Untuk k=6

Dibawah ini merupakan tabel *confussion matrix* yang dihasilkan dari aplikasi R, jika nilai k-nya 6. Selain itu pula dilakukan penjumlahan dari masing-masing baris dan kolom.

Tabel 5.13 *Confussion* Matrix dengan k=6

Aktual	Prediksi					
0	117	14	15	1	0	147
1	32	153	17	1	0	203
2	15	24	105	3	0	147
3	4	3	5	19	0	31
4	2	1	1	2	4	10
Sum	170	195	143	26	4	538

Dapat dilihat pada tabel 5.13 diatas, ketika status curah hujan aktual 0 diprediksi 0 terdapat 117 kelas, aktual 1 diprediksi 1 terdapat 153 kelas, aktual 2 diprediksi 2 terdapat 105 kelas, aktual 3 diprediksi 3 terdapat 19 kelas dan aktual 4 diprediksi 4 terdapat 4 kelas.

Selanjutnya didapatkan hasil ketepatan klasifikasi yang dibuktikan juga dengan pengerjaan manual pada Ms. Excel. Berikut ini merupakan hasilnya.

Tabel 5.14 *Confussion* Matrix k=6

Aktual	Prediksi				
	0	1	2	3	4
0	0.217472	0.026022	0.027881	0.001859	0
1	0.05948	0.284387	0.031599	0.001859	0
2	0.027881	0.04461	0.195167	0.005576	0
3	0.007435	0.005576	0.009294	0.035316	0
4	0.003717	0.001859	0.001859	0.003717	0.007435

Dalam tabel 5.14 diatas terdapat 5 kotak yang berwarna hijau, apabila kelima nilai tersebut dijumlahkan, didapatkan hasil ketepatan klasifikasi 0.739777. Untuk mendapatkan nilai kesalahan, maka 1 – 0.739777 = 0.260223 yang berarti kesalahannya sebesar 26%.

Berdasarkan keseluruhan hasil analisis yang telah dijabarkan diatas, dapat diringkas sebagai berikut.

Tabel 5.15 Summary Hasil Analisis ESKnn

	Nilai Error	Ketepatan Klasifikasi
k = 3	0.260223	0.739777
k = 4	0.2695167	0.730483
k = 5	0.2695167	0.730483
k = 6	0.260223	0.739777

Dapat dilihat pada tabel 5.15 diatas, ketika nilai k sebesar 3, maka nilai error yang didapatkan 0.260223 dan ketepatan klasifikasi yang dihasilkan sebesar 0.74 (74%). Ketika nilai k sebesar 4, maka nilai error yang didapatkan 0.2695167 dan ketepatan klasifikasi yang dihasilkan sebesar 0.73 (73%). Ketika nilai k sebesar 5, maka nilai error yang didapatkan 0.2695167 dan ketepatan klasifikasi yang dihasilkan sebesar 0.73 (73%). Ketika nilai k sebesar 6, maka nilai error yang didapatkan sebesar 0.260223 dan ketepatan klasifikasi yang dihasilkan sebesar 0.73 (74%). Berdasarkan tabel 5.15 diatas dapat dilihat pula bahwa

terdapat kesamaan antara nilai error dan ketepatan klasifikasi yang dihasilkan oleh nilai k sebesar 3 dan 6. Begitu pula dengan nilai error dan ketepatan klasifikasi yang dihasilkan oleh nilai k sebesar 4 dan 5.

Berdasarkan penjelasan diatas, nilai k yang akan peneliti pakai yaitu k = 3 dan k = 6 yang nilai ketepatan klasifikasinya sebesar 74%. Hal ini dikarenakan k 3 dan 6 memiliki ketepatan klasifikasi yang lebih besar dibandingkan k 4 dan 5. Dimana semakin besar nilai ketepatan klasifikasi, semakin bagus pula tingkat kedekatan antara nilai prediksi dan nilai aktual.

BAB VI
PENUTUP

6.1 Kesimpulan

Berdasarkan hasil penelitian dan pembahasan, dapat ditarik kesimpulan sebagai berikut :

1. Kondisi curah hujan pada tahun 2014 hingga 2018 tidak menentu. Seperti di bulan September 2014 curah hujannya sangat sedikit. Hal ini disebabkan pada bulan tersebut menurut Kepala Badan Penanggulangan Bencana Daerah (BPBD) Bogor mengatakan bahwa wilayah Bogor darurat kekeringan sejak awal hingga akhir bulan September 2014. Selain itu di bulan November 2015 curah hujannya sangat tinggi. Hal ini disebabkan pada bulan tersebut menurut Kepala Stasiun Klimatologi Dramaga Bogor mengatakan terjadi cuaca ekstrem hingga bulan Desember 2015 akibat adanya peralihan dari musim kemarau hingga musim penghujan. Namun secara keseluruhan, rata-rata curah hujan dari tahun 2014-2018 berada pada kisaran 11-16 mm.
2. Berdasarkan nilai k yang telah di gunakan diantaranya k = 3,4,5 dan 6. Nilai k yang akan di pakai yaitu k = 3 dan k = 6 yang nilai ketepatan klasifikasinya sebesar 74%. Hal ini dikarenakan k 3 dan 6 memiliki ketepatan klasifikasi yang lebih besar dibandingkan k 4 dan 5. Dimana semakin besar nilai ketepatan klasifikasi, semakin bagus pula tingkat kedekatan antara nilai prediksi dan nilai aktual. Selain itu, dilihat dari hasil *confussion matrix* untuk k = 3 dan k = 6, hasil prediksi jumlah kelas tiap status curah hujan perbedaannya cukup mencolok, dimana terdapat selisih jumlah kelas antara aktual dan prediksi. Sebagai contoh, untuk k = 3 ketika status curah hujan 0 (sangat ringan) diprediksi 0 terdapat 110 kelas dan ketika status curah hujan 3 (lebat) diprediksi 3 terdapat 24 kelas. Sedangkan untuk k = 6 ketika status curah hujan 0 (sangat ringan) diprediksi 0 terdapat 117 kelas dan ketika status curah hujan 3 (lebat) diprediksi 3 terdapat 19 kelas.

6.2 Saran

Berdasarkan hasil pembahasan serta hal-hal yang terkait dengan keterbatasan penelitian, maka terdapat beberapa hal yang perlu diperhatikan, yaitu :

1. Bagi pihak BMKG Indonesia
    a. Terkait dengan data, harus lebih diperjelas dan dilengkapi terutama masih banyak terdapat data yang kosong atau bahkan tidak ada pengukuran dalam hari tersebut. Setidaknya diberi keterangan maksud dari data kosong tersebut atau sebab tidak adanya pengukuran.
    b. Ketika login untuk mengambil data, cukup sulit dikarenakan apabila sebelumnya sudah pernah membuat akun dan akun tersebut tidak aktif beberapa lama, sudah tidak bisa login lagi dan harus membuat akun baru dengan email yang baru.
2. Bagi pemerintah, dapat menjadikan penelitian ini sebagai sarana informasi dan pengetahuan terkait dengan intensitas curah hujan dan pengklasifikasiannya di suatu daerah.
3. Bagi peneliti selanjutnya
    a. Dapat digunakan sebagai bahan perbandingan dan dan referensi untuk penelitian serta sebagai bahan pertimbangan untuk lebih memperdalam penelitian selanjutnya.
    b. Diharapkan dapat menambah variabel yang digunakan maupun menambah populasi curah hujan yang lebih luas (misal: se-Jawa Barat atau pulau Jawa) atau melakukan perbandingan dengan metode lain.
    c. Diharapkan untuk mengkaji lebih banyak sumber dan referensi yang terkait agar hasil penelitiannya lebih baik.

DAFTAR PUSTAKA

(2016). Dipetik Maret 5, 2019, dari Diskominfostandi Kota Bogor: https://kotabogor.go.id/index.php/page/detail/9/letak-geografis

Alkhatib K, N. H. (2013). Stock Price Prediction Using k-Nearest Neighbor (kNN) Algorithm. *International Journal of Business, Humanities and Technology*.

Altman, N. S. (1992, July 10). An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression. *The American Statistician*, 175-185.

Anthony. (2005). *Understanding Interobserver Agreement: The Kappa Statistic.*

Banjarsari, M. A. (2015). Penerapan K-Optimal Pada Algoritma Knn untuk Prediksi Kelulusan Tepat Waktu Mahasiswa Program Studi Ilmu Komputer Fmipa Unlam Berdasarkan IP Sampai Dengan Semester 4. *Kumpulan Jurnal Ilmu Komputer*.

BMKG. (t.thn.). Dipetik Desember 5, 2019, dari www.bmkg.go.id

BMKG. (2016, December 22). Dipetik March 12, 2020, dari https://www.bmkg.go.id/berita/?p=dua-siklon-tropis-tumbuh-dekat-wilayah-indonesia-ini-analisis-bmkg-terkait-dampaknya&lang=ID&tag=meteorologi.

Bogor, D. K. (2016). Dipetik April 3, 2019, dari https://kotabogor.go.id/index.php/page/detail/9/letak-geografis

Budiman, S. (2015). Makalah Pembelajaran Mesin KNN (K-Nearest Neighbor).

Christopher, C. (2010). *Encyclopedia Britannica: Definisi Data Mining.*

Cover T, H. P. (1967). *Nearest neighbor pattern classification.* IEEE Trans Inf Theory.

F, G. (2011). *Data Mining: Concepts, Model and Techniques.* Berlin, German: Springer.

Fajrin, A. A. (2018). Penerapan Data Mining Untuk Analisis Pola Pembelian Konsumen Dengan Algoritma FP- GROWTH Pada Data Transaksi Penjualan Spare Part Motor. *Kumpulan Jurnal Ilmu Komputer (KLIK)*.

Fauziah, S. N. (2008). Curah Hujan dan Potensi Bencana Gerakan Tanah.

Foundation, T. R. (t.thn.). Dipetik April 28, 2009, dari https://www.r-project.org/about.html

Gul, A. (2016). *Ensemble of a Subset of kNN Classifiers.*

Hornik, K. (2020). *The R FAQ: Why is R named R?*

John, F. (2005). *Using the R Statistical Computing Environment to Teach Social Statistics Courses.*

Maimon. (2010). *Web Mining dalam Data Mining and Knowledge Discovery Handbook.* Israel: Springer.

Manalu, M. T. (2016). Jaringan Syaraf Tiruan Untuk Memprediksi Curah Hujan Sumatera Utara Dengan Metode Back Propagation (Studi Kasus : BMKG Medan). *Jurnal Riset Komputer (JURIKOM)*.

Menarianti, I. (2015). Klasifikasi Data Mining Dalam Menentukan Pemberian Kredit Bagi Nasabah Koperasi. *Jurnal Ilmiah Teknosains*.

Moritz. (2018). Packages "imputeTS".

Rakhmalia, R. I. (2018). Perbandingan Hasil Metode Naive Bayes Classifier dan Support Vector Machine Dalam Klasifikasi Curah Hujan.

Ramadhani, R. D. (2019, June 20). Dipetik July 8, 2020, dari medium.com.

Sholihin, M., Fuad, N., & Khamiliyah, N. (2013). Sistem Pendukung Keputusan Penentuan Warga Penerima Jamkesmas Dengan Metode Fuzzy Tsukamoto. *Jurnal Teknika*, 501-505.

Sorjamaa A, H. J. (2005). Mutual Information and k-Nearest Neighbors Approximator for Time Series Prediction.

Steffen Moritz, T. B.-B. (2017). imputeTS: Time Series Missing Value Imputation in R.

Suwarno. (2016). Penerapan Algoritma Bayesian Regularization Backpropagation Untuk Memprediksi Penyakit Diabetes. *Jurnal MIPA*.

Vance, A. (2009). Data Analysts Captivated by R's Power.

Witten. (2011). *Data Mining: Practical Machine Learning Tools and Techniques.* Burlington, United States of America: Morgan Kaufmann.

Yu. (2007). *Application and Comparison of Classification Techniques in Controlling Credit Risk.* Singapore: World Scientific.

LAMPIRAN

Lampiran 1. Data Penelitian

No.	Tanggal	uhu rata-rata C	Kelembaban rata-rata (%)	Lama penyinaran (jam)	Arah angin saat kec. Max (deg)	Curah hujan (mm)
1	1/1/2014	25.3	89	0	360	10
2	2/1/2014	25.8	86	5.7	270	0.3
3	3/1/2014	25.6	86	5.3	315	0.9
4	4/1/2014	25.6	87	NA	45	0
5	5/1/2014	26.4	79	5.4	315	3.2
6	6/1/2014	25.7	82	2.4	270	3.3
7	7/1/2014	25.6	84	3.6	270	0
8	8/1/2014	23.5	95	0	225	1.8
9	9/1/2014	26.1	79	5.8	225	36.8
10	10/1/2014	26.6	79	7.7	315	0
…	…	…	…	…	…	…
…	…	…	…	…	…	…
…	…	…	…	…	…	…
…	…	…	…	…	…	…
1783	22/12/2018	26.6	85	8.4	315	NA
1784	23/12/2018	25.2	88	0.6	270	7.6
1785	24/12/2018	26	82	1	360	2.7
1786	25/12/2018	25.4	88	2.8	270	NA
1787	26/12/2018	24.8	90	0.4	330	22.9
1788	27/12/2018	26.1	83	0	360	1
1789	28/12/2018	27.3	76	4.3	257	1.8
1790	29/12/2018	27.1	78	7.5	223	NA
1791	30/12/2018	28	73	5	197	NA
1792	31/12/2018	25.7	82	8.8	230	NA

Lampiran 2. Data missing value yang telah diestimasi

No.	Tanggal	uhu rata-rata C	Kelembaban rata-rata (%)	Lama penyinaran (jam)	Arah angin saat kec. Max (deg)	Curah hujan (mm)
1	1/1/2014	25.3	89	0	360	10
2	2/1/2014	25.8	86	5.7	270	0.3
3	3/1/2014	25.6	86	5.3	315	0.9
4	4/1/2014	25.6	87	5.35	45	0
5	5/1/2014	26.4	79	5.4	315	3.2
6	6/1/2014	25.7	82	2.4	270	3.3
7	7/1/2014	25.6	84	3.6	270	0
8	8/1/2014	23.5	95	0	225	1.8
9	9/1/2014	26.1	79	5.8	225	36.8
10	10/1/2014	26.6	79	7.7	315	0
…	…	…	…	…	…	…
…	…	…	…	…	…	…
…	…	…	…	…	…	…
…	…	…	…	…	…	…
1783	22/12/2018	26.6	85	8.4	315	6.6
1784	23/12/2018	25.2	88	0.6	270	7.6
1785	24/12/2018	26	82	1	360	2.7
1786	25/12/2018	25.4	88	2.8	270	12.8
1787	26/12/2018	24.8	90	0.4	330	22.9
1788	27/12/2018	26.1	83	0	360	1
1789	28/12/2018	27.3	76	4.3	257	1.8
1790	29/12/2018	27.1	78	7.5	223	1.8
1791	30/12/2018	28	73	5	197	1.8
1792	31/12/2018	25.7	82	8.8	230	1.8

Lampiran 3. Pengkategorian Data Curah Hujan

No.	Tanggal	Curah hujan (mm)	Label	Status
1	1/1/2014	10	1	Ringan
2	2/1/2014	0.3	0	Sangat Ringan
3	3/1/2014	0.9	0	Sangat Ringan
4	4/1/2014	0	0	Sangat Ringan
5	5/1/2014	3.2	0	Sangat Ringan
6	6/1/2014	3.3	0	Sangat Ringan
7	7/1/2014	0	0	Sangat Ringan
8	8/1/2014	1.8	0	Sangat Ringan
9	9/1/2014	36.8	2	Sedang
10	10/1/2014	0	0	Sangat Ringan
…	…	…	…	…
…	…	…	…	…
…	…	…	…	…
…	…	…	…	…
1783	22/12/2018	6.6	1	Ringan
1784	23/12/2018	7.6	1	Ringan
1785	24/12/2018	2.7	0	Sangat Ringan
1786	25/12/2018	12.8	1	Ringan
1787	26/12/2018	22.9	2	Sedang
1788	27/12/2018	1	0	Sangat Ringan
1789	28/12/2018	1.8	0	Sangat Ringan
1790	29/12/2018	1.8	0	Sangat Ringan
1791	30/12/2018	1.8	0	Sangat Ringan
1792	31/12/2018	1.8	0	Sangat Ringan

Lampiran 4. Syntax

```
#Menghilangkan Data Missing
library(imputeTS)
aa = read.delim("clipboard")
no.miss <- na_interpolation(aa)
no.miss
is.na(aa)
sum(is.na(aa))
summary(aa)
write.csv(no.miss, file = "nomiss1.csv")

#ESKNN
library(ESKNN)
data<-read.delim("clipboard")
data

# Spliting the data into testing and training parts.
Class <- data[,names(data)=="Label"]
Class
data$Label<-as.factor(Class)
data$Label
set.seed(123)
train <- data[sample(1:nrow(data),0.7*nrow(data)),]
train
test <- data[-(sample(1:nrow(data),0.7*nrow(data))),]
test
ytrain<-train[,names(train)=="Label"]
ytrain
xtrain<-train[,names(train)!="Label"]
xtrain
xtest<-test[,names(test)!="Label"]
xtest
ytest <- test[,names(test)=="Label"]
ytest

# Trian esknnClass using training data
model<-esknnClass(xtrain, ytrain,k=3)
model
# Predict on test data
resClass<-Predict.esknnClass(model,xtest,ytest,k=3)
# Returning Objects are predicted class labels, confusion matrix and classification
error
resClass
resClass$predClass
resClass$ConfMatrix
resClass$ClassError
```

Lampiran 5. Output

k=3

```
> resClass<-Predict.esknnClass(model,xtest,ytest,k=3)
> # Returning Objects are predicted class labels, confusion matrix and class
ification error
> resClass
$PredClass
  [1] 2 1 1 1 4 5 2 2 2 1 3 3 2 1 1 3 3 2 3 3 2 3 2 1 2 1 2 3 5 1 1 4 2 3
 [35] 1 1 3 1 1 2 2 1 1 1 1 1 2 2 2 1 1 1 1 2 1 1 3 1 1 2 2 5 1 1 1 3 1 1
 [69] 1 1 1 1 1 3 2 1 2 3 3 1 1 2 1 1 3 1 4 3 5 2 1 2 1 1 1 4 3 1 1 1 2 1
[103] 1 1 2 1 2 4 2 2 3 2 1 2 1 2 1 2 1 1 2 1 1 1 2 2 2 3 2 3 2 3 2 2 2 3
[137] 3 3 2 3 2 1 2 1 3 3 1 1 1 1 1 1 1 1 1 1 1 1 3 1 1 1 1 1 2 1 1 3 1 1
[171] 1 1 1 3 4 3 3 2 1 2 1 4 3 3 1 1 1 3 1 2 2 2 4 1 1 1 2 3 4 1 4 2 4 3
[205] 4 1 1 2 2 3 3 4 2 1 3 3 2 1 1 4 2 2 3 1 1 3 1 2 2 3 4 4 2 4 4 1 3 1
[239] 2 2 4 3 3 2 3 1 2 4 3 1 3 2 3 4 1 2 2 2 4 3 2 2 2 1 2 2 3 3 2 1 2 2
[273] 2 2 2 2 3 3 3 2 2 2 1 1 3 4 2 3 3 2 5 2 3 2 2 2 2 1 1 1 3 3 2 3 2 1
[307] 1 4 1 2 2 3 3 2 3 2 2 3 2 1 3 2 2 2 2 1 2 3 3 3 4 2 2 2 3 2 2 4 1 1
[341] 2 2 2 2 3 2 2 2 3 4 1 3 3 3 2 3 2 3 2 2 2 2 2 4 1 3 3 4 1 1 4 3 4 2
[375] 2 2 1 3 3 3 3 4 3 2 3 2 2 2 3 3 3 2 2 3 2 2 2 3 3 1 2 1 2 3 2 1 3 3
[409] 2 1 2 1 1 3 1 3 3 4 1 2 2 2 1 2 1 3 3 3 3 1 3 2 3 2 2 2 4 3 3 3 2 3
[443] 3 2 2 1 1 2 2 2 2 3 1 2 3 2 3 3 3 3 3 3 3 2 3 2 3 2 3 2 2 2 3 5 2
[477] 2 3 1 2 3 4 2 2 1 3 3 2 2 2 2 2 2 2 1 2 2 3 2 3 2 1 3 3 3 2 2 1 3 4
[511] 1 1 1 3 2 2 2 4 2 2 2 2 2 4 1 2 2 3 2 3 1 2 2 2 2 3 1 1

$ConfMatrix
          Predicted.Class
True.Class   0   1   2   3   4
         0 110  15  19   3   0
         1  26 156  18   3   0
         2  16  22 103   6   0
         3   2   3   1  24   1
         4   2   1   0   2   5

$ClassError
[1] 0.260223
```

```
> resClass$predClass
NULL
> resClass$ConfMatrix
          Predicted.Class
True.Class   0   1   2   3   4
         0 110  15  19   3   0
         1  26 156  18   3   0
         2  16  22 103   6   0
         3   2   3   1  24   1
         4   2   1   0   2   5
> resClass$ClassError
[1] 0.260223
```

k=4

```
> resClass<-Predict.esknnClass(model,xtest,ytest,k=4)
> # Returning Objects are predicted class labels, confusion matrix and class
ification error
> resClass
$PredClass
  [1] 2 1 1 1 3 5 2 2 2 1 3 3 2 1 1 3 1 2 3 3 2 3 2 1 2 1 2 1 5 1 1 4 2 3
 [35] 1 1 3 1 1 2 2 1 1 1 3 1 2 2 2 1 1 1 1 2 1 1 3 1 1 2 2 4 1 1 1 3 1 1
 [69] 1 1 1 1 1 3 2 1 1 3 3 1 1 2 1 1 3 1 4 3 5 3 1 2 1 1 1 4 3 1 1 1 2 1
[103] 1 1 2 1 2 3 2 2 3 2 1 2 1 2 1 2 1 1 2 1 1 1 3 2 2 3 2 3 2 3 2 2 2 3
[137] 3 3 2 3 2 1 2 1 1 3 1 1 1 1 1 1 1 1 1 1 1 3 1 1 1 1 1 1 1 1 3 1 1
[171] 1 1 1 3 4 3 3 2 1 2 1 1 3 3 1 1 1 3 1 2 2 2 3 1 1 1 2 3 4 1 3 2 3 3
[205] 4 1 1 2 2 3 3 4 1 1 3 3 2 1 1 4 2 2 3 1 1 3 2 2 2 3 4 4 2 4 4 1 3 1
[239] 2 2 4 3 3 2 3 1 2 4 3 1 3 2 3 4 1 2 2 2 4 3 2 2 2 1 2 2 3 3 2 1 2 2
[273] 2 2 2 2 3 3 3 2 4 2 1 1 1 4 2 3 3 2 5 2 1 2 4 2 2 1 1 1 3 3 2 3 2 1
[307] 1 4 1 2 2 3 3 2 2 2 2 3 2 1 3 2 2 2 2 1 2 2 3 3 4 2 2 2 3 3 2 4 1 1
[341] 2 2 2 2 3 2 2 2 3 4 2 3 3 3 2 3 2 3 2 2 2 2 2 4 1 3 3 4 1 1 4 1 4 1
[375] 2 2 1 3 3 3 3 4 3 2 3 2 2 2 3 3 3 2 2 3 2 1 2 3 3 1 2 1 2 3 2 1 3 3
[409] 2 1 2 1 1 3 1 3 3 4 1 2 2 2 1 3 1 3 3 3 3 3 3 2 3 2 2 2 4 3 3 1 2 3
[443] 2 2 2 1 1 2 2 2 2 3 1 2 3 2 2 3 3 3 3 3 2 3 2 3 2 3 2 3 2 2 2 5 2
[477] 1 3 1 2 3 4 2 2 1 3 3 2 2 2 2 2 2 2 1 2 2 3 2 3 2 2 3 3 3 2 2 1 3 4
[511] 1 1 1 3 2 2 2 4 2 2 3 2 2 4 1 2 2 3 2 3 1 2 2 2 2 3 2 1

$ConfMatrix
          Predicted.Class
True.Class   0   1   2   3   4
         0 115  13  17   2   0
         1  31 152  15   5   0
         2  15  23 103   6   0
         3   2   3   6  19   1
         4   1   2   0   3   4

$ClassError
[1] 0.2695167
```

```
> resClass$predClass
NULL
> resClass$ConfMatrix
          Predicted.Class
True.Class   0   1   2   3   4
         0 115  13  17   2   0
         1  31 152  15   5   0
         2  15  23 103   6   0
         3   2   3   6  19   1
         4   1   2   0   3   4
> resClass$ClassError
[1] 0.2695167
```

k=5

```
> resClass<-Predict.esknnClass(model,xtest,ytest,k=5)
> # Returning objects are predicted class labels, confusion matrix and class
ification error
> resClass
$PredClass
  [1] 2 1 1 1 3 5 2 2 2 1 3 3 1 1 1 3 1 2 3 3 2 2 2 1 2 1 2 1 5 3 1 4 2 3
 [35] 1 1 3 1 1 2 2 1 1 1 1 1 1 2 2 1 1 1 1 2 1 1 3 2 1 2 2 4 1 1 1 3 1 1
 [69] 1 1 1 1 1 1 2 1 1 3 2 1 1 2 1 1 3 1 4 3 5 2 1 2 1 3 1 4 3 1 1 1 2 1
[103] 1 1 2 1 2 3 2 2 3 2 1 2 1 2 1 2 1 1 2 1 1 1 3 1 2 3 2 3 2 3 2 2 2 3
[137] 3 3 2 3 2 1 2 1 1 3 1 1 1 1 1 1 1 1 1 1 1 3 3 1 1 1 2 1 1 1 3 1 1
[171] 1 1 1 3 4 3 3 2 1 2 1 1 3 3 1 1 1 3 1 1 2 2 3 1 1 1 2 3 4 2 3 2 3 3
[205] 4 1 1 2 2 3 3 4 1 2 3 3 2 1 1 4 2 2 3 1 1 3 1 2 2 3 4 4 2 4 4 1 3 2
[239] 2 2 4 3 3 2 3 1 2 4 3 1 3 2 3 4 1 2 2 2 4 2 2 2 2 1 2 2 3 3 2 1 2 2
[273] 2 2 2 2 3 3 3 2 2 2 1 1 1 4 2 3 3 2 4 2 3 2 2 2 2 2 2 1 3 3 2 3 2 1
[307] 1 4 2 2 2 3 3 2 2 2 2 3 2 1 1 2 2 2 2 2 2 2 3 3 4 2 2 2 3 2 2 4 1 1
[341] 2 2 2 2 3 2 2 3 3 4 2 3 3 3 4 3 2 3 2 2 2 2 2 1 1 3 3 4 1 2 4 1 4 1
[375] 2 1 1 3 3 3 3 4 3 2 1 2 2 2 3 3 3 2 2 3 2 1 2 3 3 1 2 1 2 3 2 1 3 3
[409] 2 1 2 1 1 3 1 3 2 3 1 2 2 3 1 3 1 2 3 3 3 3 3 2 3 2 2 2 4 3 3 1 2 3
[443] 2 2 2 1 1 2 2 3 1 3 1 2 3 2 2 3 3 2 3 3 3 2 3 2 3 2 3 2 3 2 2 2 5 2
[477] 1 3 1 2 3 3 2 2 1 3 3 2 2 2 2 2 2 2 2 2 2 2 2 3 2 2 3 3 3 1 2 1 3 4
[511] 1 1 2 3 2 2 2 1 2 2 2 2 2 4 1 2 2 3 2 3 1 1 2 2 2 3 1 1

$ConfMatrix
          Predicted.Class
True.Class   0   1   2   3   4
         0 112  17  16   2   0
         1  31 157  13   2   0
         2  15  26 101   5   0
         3   4   2   6  19   0
         4   2   1   0   3   4

$ClassError
[1] 0.2695167
```

```
> resClass$predClass
NULL
> resClass$ConfMatrix
          Predicted.Class
True.Class   0   1   2   3   4
         0 112  17  16   2   0
         1  31 157  13   2   0
         2  15  26 101   5   0
         3   4   2   6  19   0
         4   2   1   0   3   4
> resClass$ClassError
[1] 0.2695167
```

k=6

```
> resClass<-Predict.esknnClass(model,xtest,ytest,k=6)
> # Returning Objects are predicted class labels, confusion matrix and class
ification error
> resClass
$PredClass
  [1] 2 1 1 1 4 5 2 2 2 1 3 3 2 1 1 3 3 2 3 3 2 2 2 1 2 1 2 1 5 3 1 4 2 3
 [35] 1 1 3 1 1 2 2 1 1 1 1 1 1 2 2 1 2 1 1 2 1 1 3 1 1 2 2 4 1 1 1 3 1 1
 [69] 1 1 1 1 1 1 2 1 1 3 1 1 1 2 1 1 2 1 4 3 5 2 1 2 1 3 1 1 3 1 1 1 2 1
[103] 1 1 2 1 2 3 2 2 3 2 1 2 1 2 1 1 1 1 2 1 1 1 3 1 1 3 2 3 2 3 2 2 2 3
[137] 3 3 2 3 2 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 3 3 1 1 1 1 1 1 1 3 1 1
[171] 1 1 1 3 3 3 3 2 1 2 1 1 3 3 1 1 1 3 1 1 2 2 3 1 1 1 2 3 4 2 4 2 3 3
[205] 4 1 1 2 2 3 3 3 1 2 3 3 2 1 1 3 2 2 3 1 1 3 1 2 2 3 4 4 2 4 4 1 3 2
[239] 2 2 4 3 3 2 3 1 2 4 3 1 3 2 3 4 1 2 2 2 4 2 2 2 2 1 2 2 3 3 2 1 2 2
[273] 2 2 2 2 3 3 3 2 2 2 1 1 3 4 2 3 3 2 4 2 3 2 2 2 2 2 2 1 3 3 2 3 2 1
[307] 1 3 1 2 2 3 3 2 2 2 2 3 2 1 1 3 2 2 2 2 2 3 3 3 4 2 2 2 3 2 2 3 1 1
[341] 2 1 2 2 3 2 2 2 3 3 2 3 3 3 4 3 2 3 2 2 2 2 1 1 1 3 3 3 2 1 4 3 4 1
[375] 2 1 1 3 3 3 3 4 3 1 1 2 2 2 3 3 3 2 2 3 2 1 2 3 2 1 2 1 2 3 2 1 3 3
[409] 3 1 2 1 1 3 1 3 2 4 1 2 2 3 1 3 1 3 3 3 3 3 3 2 3 2 2 2 4 3 3 1 2 3
[443] 2 2 2 1 1 2 2 2 1 3 1 2 3 2 2 3 3 3 3 3 3 2 3 2 3 2 3 2 3 2 3 2 5 2
[477] 1 3 1 2 3 3 2 2 1 3 3 2 2 2 2 2 2 2 2 2 2 2 2 3 2 2 3 3 3 1 2 1 3 4
[511] 1 1 2 3 2 2 2 1 2 2 2 2 2 2 1 2 2 3 2 4 1 1 2 2 2 3 1 1

$ConfMatrix
          Predicted.Class
True.Class   0   1   2   3   4
         0 117  14  15   1   0
         1  32 153  17   1   0
         2  15  24 105   3   0
         3   4   3   5  19   0
         4   2   1   1   2   4

$ClassError
[1] 0.260223
```

```
> resClass$predClass
NULL
> resClass$ConfMatrix
          Predicted.Class
True.Class   0   1   2   3   4
         0 117  14  15   1   0
         1  32 153  17   1   0
         2  15  24 105   3   0
         3   4   3   5  19   0
         4   2   1   1   2   4
> resClass$ClassError
[1] 0.260223
```