

***K-MEANS CLUSTERING* UNTUK KATEGORISASI CUACA
(STUDI KASUS: CUACA DI WILAYAH BANDUNG,
SEMARANG, DAN SURABAYA)**

SKRIPSI

Untuk memenuhi sebagian persyaratan

mencapai derajat Sarjana S-1

Program Studi Teknik Informatika

Disusun Oleh

ANDRI PUJI PRASETIYO

15650033

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

Kepada

PROGRAM STUDI TEKNIK INFORMATIKA

FAKULTAS SAINS DAN TEKNOLOGI

UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA

YOGYAKARTA

2019



KEMENTERIAN AGAMA
UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA
FAKULTAS SAINS DAN TEKNOLOGI
Jl. Marsda Adisucipto Telp. (0274) 540971 Fax. (0274) 519739 Yogyakarta 55281

PENGESAHAN TUGAS AKHIR

Nomor : B-3028/Un.02/DST/PP.00.9/08/2019

Tugas Akhir dengan judul : K-MEANS CLUSTERING UNTUK KATEGORISASI CUACA (STUDI KASUS : CUACA DI WILAYAH BANDUNG, SEMARANG DAN SURABAYA)

yang dipersiapkan dan disusun oleh:

Nama : ANDRI PUJI PRASETIYO
Nomor Induk Mahasiswa : 15650033
Telah diujikan pada : Senin, 22 Juli 2019
Nilai ujian Tugas Akhir : A-

dinyatakan telah diterima oleh Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta

TIM UJIAN TUGAS AKHIR

Ketua Sidang

Muhammad Didik Rohmad Wahyudi, S.T., MT.
NIP. 19760812 200901 1 015

Penguji I

Dr. Shofwatul Uyun, S.T., M.Kom.
NIP. 19820511 200604 2 002

Penguji II

Dr. Agung Fatwanto, S.Si., M.Kom.
NIP. 19770103 200501 1 003

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

Yogyakarta, 22 Juli 2019

UIN Sunan Kalijaga

Fakultas Sains dan Teknologi

Dekan



Dr. Agung Fatwanto, S.Si., M.Kom.
NIP. 19770103 200501 1 003



SURAT PERSETUJUAN SKRIPSI/TUGAS AKHIR

Hal : Permohonan

Lamp :

Kepada

Yth. Dekan Fakultas Sains dan Teknologi

UIN Sunan Kalijaga Yogyakarta

di Yogyakarta

Assalamu'alaikum wr. wb.

Setelah membaca, meneliti, memberikan petunjuk dan mengoreksi serta mengadakan perbaikan seperlunya, maka kami selaku pembimbing berpendapat bahwa skripsi Saudara:

Nama : Andri Puji Prasetyo

NIM : 15650033

Judul Skripsi : **K-Means Clustering Untuk Kategorisasi Cuaca (Studi Kasus: Cuaca di Wilayah Bandung, Semarang, dan Surabaya)**

sudah dapat diajukan kembali kepada Program Studi Teknik Informatika Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta sebagai salah satu syarat untuk memperoleh gelar Sarjana Strata Satu dalam Program Studi Teknik Informatika

Dengan ini kami berharap agar skripsi/tugas akhir Saudara tersebut di atas dapat segera dimunaqsyahkan. Atas perhatiannya kami ucapkan terima kasih.

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

Wassalamu'alaikum wr. wb.

Yogyakarta, 19 Juli 2019

Pembimbing

Muhammad Didik Rohmad Wahyudi, S.T., MT.

NIP. 19760812 200901 1 015

PERNYATAAN KEASLIAN SKRIPSI

Saya yang bertanda tangan dibawah ini:

Nama : Andri Puji Prasetyo

NIM : 15650033

Jurusan : Teknik Informatika

Fakultas : Sains dan Teknologi

Menyatakan bahwa skripsi yang berjudul "*K-Means Clustering* Untuk Kategorisasi Cuaca (Studi Kasus: Cuaca di Wilayah Bandung, Semarang, dan Surabaya)" tidak terdapat pada karya yang pernah di ajukan untuk memperoleh gelar sarjana di suatu Perguruan Tinggi, dan sepengetahuan penulis tidak terdapat karya atau pendapat yang pernah ditulis oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan di sebutkan dalam daftar pustaka.

Yogyakarta, 19 Juli 2019

Yang menyatakan



Andri Puji Prasetyo
NIM. 15650033

KATA PENGANTAR

Puji syukur kehadiran Allah SWT yang telah melimpahkan rahmat dan hidayah-Nya sehingga penulis dapat menyelesaikan penelitian yang berjudul *K-Means Clustering* Untuk Kategorisasi Cuaca (Studi Kasus: Cuaca di Wilayah Bandung, Semarang dan Surabaya) sebagai salah satu syarat untuk mencapai gelar sarjana program studi Teknik Informatika UIN Sunan Kalijaga Yogyakarta. Sholawat serta salam selalu tercurahkan kepada junjungan kita Nabi Agung Muhammad SAW beserta seluruh keluarga dan sahabat beliau.

Penulis menyadari bahwa apa yang dilakukan dalam penyusunan laporan penelitian ini masih terlalu jauh dari kesempurnaan. Oleh karena itu, penulis mengharapkan kritik dan saran yang berguna dalam penyempurnaan analisa ini di masa yang akan datang. Semoga apa yang telah penulis lakukan dapat bermanfaat bagi pembaca.

Tidak lupa penulis juga mengucapkan terimakasih kepada pihak-pihak yang telah membantu dalam penyelesaian skripsi ini, baik secara langsung maupun tidak langsung. Ucapan terimakasih penulis sampaikan kepada:

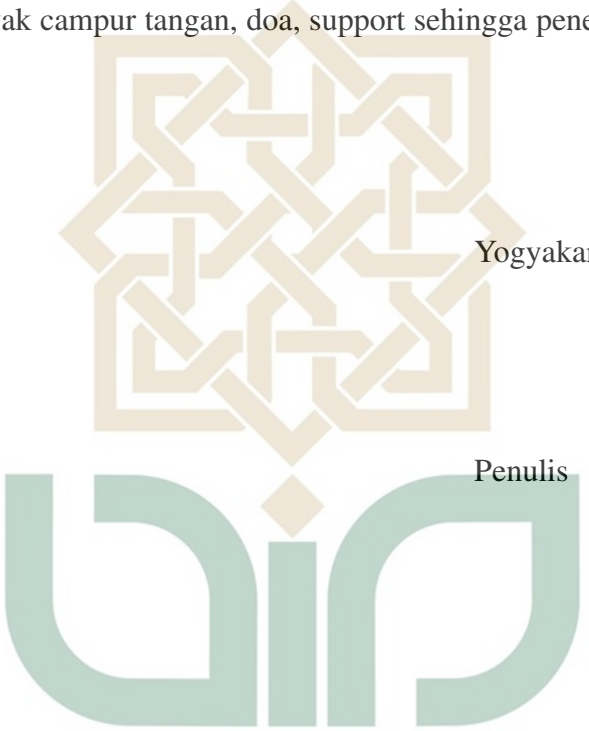
- 1 Bapak Prof. Drs. K.H. Yudian Wahyudi, M.A., Ph.D., selaku Rektor UIN Sunan Kalijaga Yogyakarta.
- 2 Bapak Dr. Murtono, M.Si., selaku Dekan Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta.
- 3 Bapak Sumarsono, S.T., M.Kom., selaku Ketua Program Studi Teknik Informatika UIN Sunan Kalijaga Yogyakarta.

- 4 Bapak Dr. Bambang Sugiantoro, M.T., selaku Dosen Pembimbing Akademik.
- 5 Bapak Muhammad Didik Rohmad Wahyudi, S.T., MT., selaku dosen pembimbing skripsi yang telah sabar dan meluangkan waktunya untuk memberikan motivasi, koreksi dan kritik saran kepada penulis sehingga skripsi ini dapat diselesaikan.
- 6 Dr. Agung Fatwanto, S.Si., M.Kom., Agus Mulyanto, S.Si., M.Kom., Ph.D., Aulia Faqih Rifai, M.Kom., M. Taufiq NUruzzaman, S.T., Maria Ulfah Siregar, S.Kom. MIT., Ph.D., Nurochman, S.Kom., M.Kom., Rahmat Hidayat, S.Kom., M.Cs., Dr. Shofwatul Uyun, S.T., M.Kom., selaku dosen pengampu mata kuliah program studi Teknik Informatika UIN Sunan Kalijaga Yogyakarta yang telah banyak membantu penulis hingga penulis dapat menyusun tugas akhir.
- 7 Seluruh staf dan karyawan Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta.
- 8 Ucapan terimakasih yang terdalam untuk kedua orangtua, Ibu Sudarti dan Bapak Sumanto Rasbin yang selalu memberikan doa, perhatian, kasih sayang dan semua support yang telah diberikan.
- 9 Ucapan terimakasih yang terdalam untuk kedua saudara saya, Mas Heri Purwadi dan Ayla Nurasyza yang selalu memberikan motivasi dan semangat sepanjang pengerjaan skripsi.
- 10 Teman-teman begadang mengerjakan skripsi Mr. Fitree Tahe, Lana Rahim, Tesya Nurintan, Khoirul Umam serta Nafi yang sudah meminjamkan buku yang sangat membantu sekali dalam pengerjaan skripsi.

- 11 Seluruh Teman-Teman Teknik Informatika 2015 yang tidak dapat penulis sebutkan satu-satu.
- 12 Teman-teman KKN Angkatan 96 Kelompok 258.
- 13 Serta semua pihak yang tidak dapat penulis sebutkan satu persatu dan telah memberikan banyak campur tangan, doa, support sehingga penelitian ini dapat terselesaikan.

Yogyakarta, 16 Juli 2019

Penulis



STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

HALAMAN PERSEMBAHAN

Dengan penuh kasih sayang dan rasa syukur kepada Allah SWT, Karya sederhana ini saya persembahkan untuk :

”Kedua Orang Tua”



HALAMAN MOTO

”Dan Kami perintahkan kepada manusia (berbuat baik) kepada dua orang ibu-bapaknya; ibunya telah mengandungnya dalam keadaan lemah yang bertambah-tambah, dan menyapihnya dalam dua tahun . Bersyukurlah kepada-Ku dan kepada dua orang ibu bapakmu, hanya kepada-Kulah kembalimu.”

(Qs. Luqman : 14)



STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

DAFTAR ISI

HALAMAN JUDUL	1
LEMBAR PENGESAHAN SKRIPSI/TUGAS AKHIR	2
LEMBAR PERSETUJUAN SKRIPSI/TUGAS AKHIR	3
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	4
KATA PENGANTAR	5
HALAMAN PERSEMBAHAN	8
HALAMAN MOTTO	9
DAFTAR ISI	10
DAFTAR TABEL	13
DAFTAR GAMBAR	15
DAFTAR LAMBANG	16
INTISARI	17
ABSTRACT	18
I PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	3
1.3. Batasan Masalah	3
1.4. Tujuan Penelitian	3
1.5. Manfaat Penelitian	3
1.6. Keaslian Penelitian	4
1.7. Sistematika Penulisan	4
II TINJAUAN PUSTAKA DAN LANDASAN TEORI	6

	11
2.1. Tinjauan Pustaka	6
2.2. Landasan Teori	12
2.2.1. Data Mining	12
2.2.2. Tahap Data Mining	13
2.2.3. <i>Unsupervised Learning</i>	15
2.2.4. Klasterisasi	16
2.2.5. K-Means Clustering	16
2.2.6. Sihoutte Coefficient	18
2.2.7. Cuaca	20
2.2.8. Python	21
III METODE PENELITIAN	23
3.1. Metode Penelitian	23
3.2. Tahap-Tahap Penelitian	23
3.2.1. Studi Pustaka	24
3.2.2. Pengumpulan Data	24
3.2.3. <i>Preprocessing</i>	24
3.2.4. Klasterisasi	24
3.2.5. Analisis dan Evaluasi	25
3.2.6. Visualisasi	25
3.2.7. Kebutuhan Sistem	25
IV HASIL DAN PEMBAHASAN	27
4.1. Pengumpulan Data	27
4.2. <i>Preprocessing</i>	28

	12
4.2.1. Pembersihan Data	28
4.2.2. <i>StandardScaler</i>	29
4.3. Klasterisasi Metode <i>K-Means Clustering</i>	31
4.4. Analisis dan Evaluasi	32
4.4.1. Hasil Evaluasi <i>Silhouette Coefficient</i>	32
V PENUTUP	59
5.1. Kesimpulan	59
5.2. Saran	59
DAFTAR PUSTAKA	63
LAMPIRAN	64
CURRICULUM VITAE	88

DAFTAR TABEL

2.1	Tabel Perbandingan Tinjauan Pustaka	9
2.2	Lanjutan Tabel Perbandingan Tinjauan Pustaka	10
2.3	Lanjutan Tabel Perbandingan Tinjauan Pustaka	11
4.1	Tabel cuaca wilayah Bandung Januari 2010	27
4.2	Tabel cuaca wilayah Bandung Januari 2010 yang telah dibersihkan .	29
4.3	Tabel kriteria intensitas curah hujan di wilayah Indonesia	31
4.4	Tabel Titik Pusat Klastersasi Data Cuaca Wilayah Bandung Ke- 1 .	35
4.5	Lanjutan Tabel Titik Pusat Klastersasi Data Cuaca Wilayah Bandung Ke- 1	36
4.6	Tabel Titik Pusat Klastersasi Data Cuaca Wilayah Bandung Ke- 2 .	36
4.7	Lanjutan Tabel Titik Pusat Klastersasi Data Cuaca Wilayah Bandung Ke- 2	37
4.8	Tabel Titik Pusat Klastersasi Data Cuaca Wilayah Bandung Ke- 3 .	38
4.9	Tabel Titik Pusat Klastersasi Data Cuaca Wilayah Bandung Ke- 4 .	39
4.10	Tabel Titik Pusat Klastersasi Data Cuaca Wilayah Bandung Ke- 5 .	40
4.11	Tabel Titik Pusat Klastersasi Data Cuaca Wilayah Bandung Ke- 6 .	41
4.12	Tabel Titik Pusat Klastersasi Data Cuaca Wilayah Semarang Ke- 1 .	42
4.13	Tabel Titik Pusat Klastersasi Data Cuaca Wilayah Semarang Ke- 2 .	43
4.14	Tabel Titik Pusat Klastersasi Data Cuaca Wilayah Semarang Ke- 3 .	44
4.15	Tabel Titik Pusat Klastersasi Data Cuaca Wilayah Semarang Ke- 4 .	45
4.16	Tabel Titik Pusat Klastersasi Data Cuaca Wilayah Semarang Ke- 5 .	46
4.17	Tabel Titik Pusat Klastersasi Data Cuaca Wilayah Semarang Ke- 6 .	47

4.18	Tabel Titik Pusat Klustersasi Data Cuaca Wilayah Surabaya Ke- 1	. 48
4.19	Tabel Titik Pusat Klustersasi Data Cuaca Wilayah Surabaya Ke- 2	. 49
4.20	Tabel Titik Pusat Klustersasi Data Cuaca Wilayah Surabaya Ke- 3	. 50
4.21	Tabel Titik Pusat Klustersasi Data Cuaca Wilayah Surabaya Ke- 4	. 51
4.22	Tabel Titik Pusat Klustersasi Data Cuaca Wilayah Surabaya Ke- 5	. 52
4.23	Tabel Titik Pusat Klustersasi Data Cuaca Wilayah Surabaya Ke- 6	. 53



DAFTAR GAMBAR

3.1 Tahapan Penelitian	23
4.1 Hasil <i>Preprocessing Standard Scaler Data Cuaca Wilayah Bandung</i>	30
4.2 Hasil <i>Preprocessing Standard Scaler Data Cuaca Wilayah Semarang</i>	30
4.3 Hasil <i>Preprocessing Standard Scaler Data Cuaca Wilayah Surabaya</i>	30
4.4 Hasil klasterisasi cuaca wilayah Bandung ke- 1	35
4.5 Hasil klasterisasi cuaca wilayah Bandung ke- 2	36
4.6 Hasil klasterisasi cuaca wilayah Bandung ke- 3	37
4.7 Hasil klasterisasi cuaca wilayah Bandung ke- 4	38
4.8 Hasil klasterisasi cuaca wilayah Bandung ke- 5	39
4.9 Hasil klasterisasi cuaca wilayah Bandung ke- 6	40
4.10 Hasil klasterisasi cuaca wilayah Semarang ke- 1	41
4.11 Hasil klasterisasi cuaca wilayah Semarang ke- 2	42
4.12 Hasil klasterisasi cuaca wilayah Semarang ke- 3	43
4.13 Hasil klasterisasi cuaca wilayah Semarang ke- 4	44
4.14 Hasil klasterisasi cuaca wilayah Semarang ke- 5	45
4.15 Hasil klasterisasi cuaca wilayah Semarang ke- 6	46
4.16 Hasil klasterisasi cuaca wilayah Surabaya ke- 1	48
4.17 Hasil klasterisasi cuaca wilayah Surabaya ke- 2	49
4.18 Hasil klasterisasi cuaca wilayah Surabaya ke- 3	50
4.19 Hasil klasterisasi cuaca wilayah Surabaya ke- 4	51
4.20 Hasil klasterisasi cuaca wilayah Surabaya ke- 5	52
4.21 Hasil klasterisasi cuaca wilayah Surabaya ke- 6	53

DAFTAR LAMBANG

$x \in A$	: x anggota A
$A \subseteq X$	: A himpunan bagian (<i>subset</i>) atau sama dengan X
$\mathbb{N}$	: himpunan semua asli
$\mathbb{Z}$	: himpunan semua bilangan bulat
$\mathbb{Z}^+$	: himpunan semua bilangan bulat positif
$\mathbb{R}$	: himpunan semua bilangan real
C_r^n	: r -kombinasi dari n unsur yang berbeda
■	: akhir suatu bukti
□	: akhir suatu contoh
$\rightarrow$	: menuju
$\sum_{i=1}^n a_i$	: penjumlahan $a_1 + a_2 + \cdots + a_n$
$\prod_{i=1}^n a_i$	: perkalian $a_1 \cdot a_2 \cdots a_n$
$p \Rightarrow q$	: jika p maka q
$\Leftrightarrow$	: jika dan hanya jika
$x \leftarrow a$	: nilai a dimasukkan ke x

INTISARI

K-Means Clustering Untuk Kategorisasi Cuaca (Studi Kasus: Cuaca di Wilayah Bandung, Semarang, dan Surabaya)

Oleh

Andri Puji Prasetyo

15650033

Informasi merupakan hal yang sangat berguna dalam berbagai aspek kehidupan termasuk diantaranya adalah informasi pemetaan cuaca. Hal ini karena pengetahuan tentang pemetaan cuaca dapat berguna dalam berbagai aktivitas kehidupan seperti keselamatan masyarakat, sosial ekonomi, produksi pertanian, perkebunan, perikanan, penerbangan, dan sebagainya dalam suatu daerah. Namun demikian, informasi cuaca tersebut belum disajikan secara spesifik. Oleh karena itu, perlu adanya otomatisasi pengkategorian cuaca untuk melakukan pemetaan cuaca agar informasi yang diperoleh lebih spesifik dan dapat bermanfaat bagi masyarakat suatu daerah.

Penelitian ini mengambil sampel cuaca dari data cuaca wilayah Bandung, Semarang, dan Surabaya dari tahun 2010 sampai dengan 2018 sebanyak 9.864 data untuk diklasterisasi secara otomatis menggunakan teknik *clustering*. Metode *clustering* yang digunakan adalah *K-Means Clustering* yang bertujuan untuk mencari jumlah *cluster* terbaik dari beberapa atribut cuaca. Atribut cuaca tersebut meliputi temperatur minimum, temperatur maksimum, temperatur rata-rata, kelembaban rata-rata, curah hujan, kecepatan angin maksimum, kecepatan angin rata-rata.

Berdasarkan hasil penelitian yang dilakukan pada kategorisasi cuaca di wilayah Bandung, Semarang, dan Surabaya. Disimpulkan pengelompokan menggunakan metode *K-Means Clustering* dengan penentuan jumlah *cluster* 4 dari ketiga wilayah tersebut setelah dilakukan evaluasi *silhouette coefficient* dengan jumlah k 2,3,4,5 dan 6 mendapatkan hasil jumlah *cluster* terbaik untuk klasterisasi cuaca wilayah Bandung yaitu sebanyak 4 *cluster* dengan nilai rata-rata *silhouette coefficient* sebesar 0.216555, untuk klasterisasi cuaca wilayah Semarang yaitu sebanyak 5 *cluster* dengan nilai rata-rata *silhouette coefficient* sebesar 0.243332, sedangkan untuk klasterisasi cuaca wilayah Surabaya mendapatkan sebanyak 4 *cluster* dengan nilai rata-rata *silhouette coefficient* sebesar 0.209312.

Kata Kunci : *K-Means Clustering, silhouette coefficient, Cuaca*

ABSTRACT

K-Means Clustering For Weather Categorization (Case Study: Weather in Bandung, Semarang, and Surabaya)

By

Andri Puji Prasetyo

15650033

Information is something that is very useful in various aspects of life, including among them weather mapping information. This is because knowledge about weather mapping can be useful in a variety of life activities such as community safety, socio-economic, agricultural production, plantations, fisheries, aviation, and so on in a area. However, the weather information has not been presented specifically. Therefore, it is necessary to automate the categorization of weather to do weather mapping so that the information obtained is more specific and can be beneficial to the community of a region.

This study took weather samples from the weather data of the Bandung, Semarang, and Surabaya regions from 2010 to 2018 as many as 9,864 data to be clustered automatically using clustering techniques. The clustering method used is K-Means Clustering which aims to find the best number of clusters from several weather attributes. These weather attributes include minimum temperature, maximum temperature, average temperature, average humidity, rainfall, maximum wind speed, average wind speed.

Based on the results of research conducted on weather categorization in the area of Bandung, Semarang, and Surabaya. It was concluded that the grouping used the K-Means Clustering method by determining the number of clusters of 4 from the three regions after evaluating the silhouette coefficient with the number of k 2,3,4,5 and 6 getting the best number of clusters for the Bandung area weather clustering, namely 4 clusters with the average silhouette coefficient value is 0.216555, for the Semarang region weather clustering which is as many as 5 clusters with an average silhouette coefficient value of 0.243332, while for the weather clustering in the Surabaya region get as many as 4 clusters with an average silhouette coefficient value of 0.209312.

Keywords : K-Means Clustering, silhouette coefficient, Weather

BAB I

PENDAHULUAN

1.1. Latar Belakang

Cuaca merupakan suatu fenomena alam yang berkaitan dengan kondisi udara di bumi. Setiap harinya cuaca pasti berubah tergantung kondisi udaranya. Secara teknis cuaca adalah keadaan atmosfer di suatu tempat atau daerah selama waktu tertentu (menit, jam, hari, bulan, musim, dan seterusnya).

Cuaca adalah keadaan atmosfer pada setiap waktu kesatuan. Dengan demikian kita dapat mengatakan tentang cuaca saat ini, cuaca pukul 12, cuaca nanti sore, cuaca tanggal 17 Agustus 1945, cuaca besok pagi, cuaca minggu ini, cuaca minggu depan, cuaca bulan depan, dan seterusnya apabila waktu-waktu tersebut diartikan sebagai satu satuan waktu. Jadi cuaca menyatakan keadaan yang berlangsung pada saat atau selama waktu kesatuan (Wirjohamidjoyo & Swarinoto 2010).

Cuaca di Indonesia ini diamati oleh suatu badan yaitu BMKG (Badan Meteorologi, Klimatologi, dan Geofisika). BMKG memberikan informasi yang berguna bagi berbagai penelitian tentang cuaca dan iklim di Indonesia. Informasi berupa data cuaca dari BMKG tersebut yang nantinya diperlukan untuk melakukan penelitian tentang kategorisasi cuaca di Indonesia. Data cuaca yang didapat akan diolah sedemikian rupa agar dapat dikategorisasi. Untuk menganalisis data tersebut diperlukan proses data mining.

Teknik datamining yang dapat digunakan untuk mengelompokkan data cuaca menurut kemiripan datanya adalah menggunakan teknik *clustering*. *Clustering*

merupakan suatu metode untuk mencari dan mengelompokkan data yang memiliki kemiripan karakteristik (*similarity*) antara satu data dengan data yang lain. *Clustering* merupakan salah satu metode data mining yang bersifat tanpa arahan (*unsupervised*), maksudnya metode ini diterapkan tanpa adanya latihan (*training*) dan tanpa ada guru serta tidak memerlukan target output (Bastian et al. 2009).

Peneliti tertarik untuk melakukan penelitian ini karena cuaca merupakan salah satu variabel yang menentukan kondisi iklim. Dimana, terdapat beberapa faktor yang berpengaruh langsung terhadap tipe atau variasi iklim seperti curah hujan dan kelembaban rata-rata. Informasi mengenai cuaca adalah salah satu unsur penting dan besar pengaruhnya terhadap segala macam aktifitas kehidupan seperti keselamatan masyarakat, sosial ekonomi, produksi pertanian, perkebunan, perikanan, penerbangan, dan sebagainya dalam suatu daerah (Puspitasari & Haviluddin 2016). Oleh karena itu, dengan adanya pengkategorian cuaca secara otomatis akan memberikan informasi pemetaan cuaca yang spesifik dari suatu daerah dan berguna bagi segala macam aktifitas tersebut.

Teknik klasterisasi yang digunakan untuk mengategorikan data cuaca dalam penelitian ini adalah *K-Means Clustering*. *K-Means* merupakan metode *clustering* yang cukup sederhana dan umum dalam penggunaannya. Metode *K-Means* adalah metode yang terkenal cepat dan simpel (Arthur & Vassilvitskii 2006).

Dengan demikian peneliti tertarik untuk melakukan penelitian dengan judul ***K-Means Clustering Untuk Kategorisasi Cuaca (Studi Kasus: Cuaca di Wilayah Bandung, Semarang, dan Surabaya).***

1.2. Rumusan Masalah

Berdasarkan latar belakang di atas maka rumusan masalah dalam penelitian ini adalah bagaimana mengkategorisasikan cuaca di wilayah Bandung, Semarang, dan Surabaya dengan menggunakan model klasterisasi *K-Means Clustering*.

1.3. Batasan Masalah

Batasan masalah pada penelitian ini yaitu :

1. Penerapan Data Mining yang digunakan dalam pengelompokkan cuaca dilakukan dengan algoritma *K-Means Clustering*.
2. Data yang digunakan diambil dari website <http://dataonline.bmkg.go.id>.
3. Data yang adalah data cuaca wilayah Bandung, Semarang dan, Surabaya dari tahun 2010 sampai 2018.

1.4. Tujuan Penelitian

Penelitian ini bertujuan untuk mengkategorisasikan cuaca secara otomatis dengan menggunakan model klasterisasi *K-Means Clustering* untuk mencari jumlah *cluster* terbaik dari setiap wilayah studi kasus serta menjadikan nilai dari setiap pusat klaster sebagai acuan pemetaan cuaca hujan.

1.5. Manfaat Penelitian

Manfaat penelitian ini yaitu:

1. Mengkategorikan cuaca Bandung, Semarang, Surabaya dalam jumlah data yang besar secara otomatis.

2. Bagi BMKG dan UIN Sunan Kalijaga, dapat digunakan sebagai tambahan referensi terkait penelitian tentang cuaca seiring dengan kemajuan teknologi.
3. Bagi Penulis, dapat memberikan pengetahuan dan wawasan yang baru, bahwa data tentang cuaca dapat di kategorisasi.
4. Bagi Pembaca, dengan adanya penelitian ini dapat digunakan sebagai sumber pembelajaran untuk penelitian selanjutnya serta untuk menambah wawasan.

1.6. Keaslian Penelitian

Penelitian tentang kategorisasi cuaca dengan menggunakan algoritma *K-Means Clustering* sampai saat ini sudah beberapa kali dilakukan. Namun, berdasarkan referensi dan tinjauan pustaka, penelitian Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta mengenai pengkategorian cuaca dengan menggunakan algoritma *K-Means Clustering* belum pernah dilakukan.

1.7. Sistematika Penulisan

Untuk memberikan gambaran dan kerangka yang jelas mengenai pokok bahasan setiap bab dalam penelitian ini, maka diperlukan sistematika penulisan. Berikut ini adalah gambaran sistematika penulisan pada masing-masing bab :

BAB I PENDAHULUAN

Bab *satu* pendahuluan, berisikan latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, keaslian penelitian, dan sistematika penulisan skripsi.

BAB II TINJAUAN PUSTAKA DAN LANDASAN TEORI

Bab *dua* tinjauan pustaka dan landasan teori, menjelaskan tentang tinjauan pustaka dan landasan teori yang berhubungan dengan topik yang akan di bahas dalam penelitian ini.

BAB III METODE PENELITIAN

Bab *tiga* metode penelitian, membahas mengenai metode yang dilakukan dan atau yang digunakan dalam menyelesaikan penelitian ini.

BAB IV HASIL DAN PEMBAHASAN

Bab *empat* hasil dan pembahasan, menjelaskan hasil dari penelitian yang dicapai. Dari proses pengambilan data hingga hasil yang diperoleh dari penelitian ini.

BAB V KESIMPULAN DAN SARAN

Bab *lima* penutup dan kesimpulan, menjelaskan kesimpulan yang dihasilkan serta saran yang akan diberikan berdasarkan hasil yang telah dicapai sehingga dapat digunakan sebagai bahan pertimbangan bagi pihak-pihak yang berkepentingan serta untuk perkembangan penelitian-penelitian selanjutnya.

BAB V

PENUTUP

5.1. Kesimpulan

Berdasarkan hasil penelitian yang dilakukan pada kategorisasi cuaca di wilayah Bandung, Semarang, dan Surabaya. Disimpulkan pengelompokan menggunakan metode *K-Means Clustering* dengan penentuan jumlah *cluster* 4 dari ketiga wilayah tersebut setelah dilakukan evaluasi *silhouette coefficient* dengan jumlah k 2,3,4,5 dan 6 mendapatkan hasil jumlah *cluster* terbaik untuk klasterisasi cuaca wilayah Bandung yaitu sebanyak 4 *cluster* dengan nilai rata-rata *silhouette coefficient* sebesar 0.216555, untuk klasterisasi cuaca wilayah Semarang yaitu sebanyak 5 *cluster* dengan nilai rata-rata *silhouette coefficient* sebesar 0.243332, sedangkan untuk klasterisasi cuaca wilayah Surabaya mendapatkan sebanyak 4 *cluster* dengan nilai rata-rata *silhouette coefficient* sebesar 0.209312.

5.2. Saran

Pada penelitian ini tentunya tidak terlepas dari kekurangan dan kelemahan. Oleh karena itu, peneliti menyarankan adanya pengembangan terhadap penelitian ini berikutnya :

1. Penelitian selanjutnya dapat menggunakan metode klasterisasi lainnya atau metode *machine learning* yang berbeda.
2. Penelitian selanjutnya dapat menggunakan ataupun menambahkan atribut selain curah hujan dan kelembaban untuk melakukan pengkategorian cuaca.

3. Diharapkan pada penelitian selanjutnya dapat mengkategorikan sekaligus memetakan cuaca berdasarkan waktu supaya mendapatkan informasi tahunan tentang pergeseran pusat klaster dari tahun ke tahun.



DAFTAR PUSTAKA

Arthur, D. & Vassilvitskii, S. (2006), How slow is the k -means method?, in 'cs.duke.edu', p. 144.

URL: <https://www.cs.duke.edu/courses/spring07/cps296.2/papers/kMeans-socg.pdf>

Bastian, A., Sujadi, H., Febrianto, G., Studi, P., Informatika, T., Majalengka, U., Universitas, J. & No, M. (2009), 'PENERAPAN ALGORITMA', *jsi.cs.ui.ac.id* (1), 26–32.

URL: <http://jsi.cs.ui.ac.id/index.php/jsi/article/view/566>

Chakraborty, S., Nagwani, N. & Dey, L. (2014), 'Weather forecasting using incremental k-means clustering', *arXiv preprint arXiv:1406.4756* .

Han, J., Pei, J. & Kamber, M. (2011), *Data mining: concepts and techniques*, Elsevier.

Handoyo, R., Rumani, R. & Nasution, S. M. (2014), 'Perbandingan Metode Clustering Menggunakan Metode Single Linkage Dan K-Means Pada Pengelompokan Dokumen', *JSM STMIK Mikroskil* 15(2), 73–82.

URL: <https://mikroskil.ac.id/ejurnal/index.php/jsm/article/view/161>

Lutz, M. (2010), *Programming Python: powerful object-oriented programming*, "O'Reilly Media, Inc."

Maimon, O. & Rokach, L. (2005), 'Data mining and knowledge discovery handbook'.

- Mujiasih, S. (2011), 'Penerapan Data Mining Untuk Prakiraan Cuaca', *Meteorologi Dan Geofisika* **12**(2), 189–195.
- Novandya, A. & Oktria, I. (2017), 'Penerapan algoritma klasifikasi data mining c4.5 pada dataset cuaca wilayah bekasi', *Format* **6**(2), 98–106.
- Novandya, Adhika., Oktria, I. (2017), 'Penerapan Algoritma Klasifikasi Data Mining C4.5 Pada Dataset Cuaca Wilayah Bekasi', *Jurnal Format* **6**(2), 98–106.
URL: <https://www.neliti.com/publications/224664/penerapan-algoritma-klasifikasi-data-mining-c45-pada-dataset-cuaca-wilayah-bekas>
- Pramesti, D. F., Furqon, M. T. & Dewi, C. (2017), 'Implementasi metode k-medoids clustering untuk pengelompokan data potensi kebakaran hutan/lahan berdasarkan peersebaran titik panas (hotspot)', *Malang: Universitas Brawijaya. ISSN: e-ISSN* .
- Puspitasari, N. & Havaluddin, H. (2016), Penerapan metode k-means dalam pengelompokan curah hujan di kalimantan timur.
- Ramadhani, Ahmad and Farmadi, Andi and Budiman, I. (2015), 'Clustering Data Cuaca Untuk Pengenalan Pola Perioditas Iklim Wilayah Pelabuhan Dengan Metode Fuzzy C-Means', *Jurnal Teknologi dan Industri* **3**, 57–64.
- Reza Noviansyah, M., Rismawan, T., Marisa Midyanti, D., Sistem Komputer, J. & MIPA Universitas Tanjungpura Jl Hadari Nawawi, F. H. (2018), PENERAPAN DATA MINING MENGGUNAKAN METODE K-NEAREST NEIGHBOR UNTUK KLASIFIKASI INDEKS CUACA KEBAKARAN BERDASARKAN

DATA AWS (AUTOMATIC WEATHER STATION) (STUDI KASUS: KABUPATEN KUBU RAYA), Technical Report 2.

URL: <http://jurnal.untan.ac.id/index.php/jcskommipa/article/view/26672>

Santosa, B. (2007), 'Data mining teknik pemanfaatan data untuk keperluan bisnis', *Yogyakarta: Graha Ilmu* **978(979)**, 756.

Subhan, A. & Fanani, A. Z. (2016), 'Penerapan Data Mining Untuk Menentukan Potensi Naive Bayes', *mahasiswa.dinus.ac.id* pp. 1–6.

URL: <http://mahasiswa.dinus.ac.id/docs/skripsi/jurnal/19417.pdf>

Suyanto, D. et al. (2017), 'Data mining untuk klasifikasi dan klasterisasi data', *Bandung: Informatika Bandung*.

Tan, P.-N., Steinbach, M. & Kumar, V. (2005), Association analysis: basic concepts and algorithms, in 'Introduction to Data mining', Vol. 321321367, Addison-Wesley Boston, MA.

Tan, P.-N., Steinbach, M. & Kumar, V. (2006), *An introduction to data mining: solution Manual*.

Wirjohamidjoyo, S. & Swarinoto, Y. (2010), *Iklim Kawasan Indonesia (Dari Aspek Dinamik - Sinoptik)*.