

PENERAPAN METODE *K-NEAREST NEIGHBOR* (KNN) UNTUK KLASIFIKASI CUACA DI WILAYAH TANJUNGPINANG

Syarifaturrahmah¹, Alena Uperiati², Eka Suswaini³

Syarifaturrahmah060398@gmail.com

Program studi Teknik Informatika, Fakultas Teknik, Universitas Maritim Raja Ali Haji

Abstract

The weather in Tanjungpinang area can be known by classification using previous data to study weather patterns that occur. Classification can be applied using the K-Nearest Neighbor method. K-Nearest Neighbor is a supervised learning algorithm. The purpose of this algorithm is to classify new objects obtained based on attributes and samples from training data. In this study KNN was used with parameters of temperature, humidity, pressure, and wind speed obtained from The Meteorological Station of Raja Haji Fisabilillah Tanjungpinang. Data taken from January 2016-December 2019. The result of this study is a classification consisting of weather types namely Thunderstorm, Lightning, Mist, Smoke, Cloudy, Prec in Sight, Rain, Haze, Shower, Lightning Rain. The results of this study obtained the highest accuracy of 70% of $k = 31, 33, 35, 37, 39, 41, 45, 47, 50$ and the lowest accuracy of 30% of $k = 7$

Keywords : *Weather, Classification, K-Nearest Neighbor.*

I. Pendahuluan

Indonesia terletak di **6° LU (Lintang Utara) – 11° LS (Lintang Selatan)** dan **95° BT (Bujur Timur) – 141° BT (Bujur Timur)** yang mana merupakan daerah tropis dengan memiliki dua musim yaitu kemarau dan hujan Nurhadi, *et al.* (2009). Diera modern ini sudah banyak teknologi berkembang yang dapat membantu kegiatan masyarakat sehari-hari misalkan prakiraan cuaca. Prakiraan cuaca diperlukan sebagai penunjang kegiatan yang akan dilakukan di luar ruangan.

Prakiraan dan prediksi cuaca memiliki fungsi untuk mengantisipasi masyarakat dalam mengalami pergantian musim di waktu yang akan datang. Prakiraan cuaca memiliki beberapa manfaat dari berbagai bidang diantaranya adalah dapat mengetahui perkiraan waktu dan juga untuk bercocok tanaman yang sesuai dengan musimnya, membantu merencanakan kegiatan yang akan dilakukan sehari-hari, serta menjaga keselamatan dalam transportasi.

Cuaca di wilayah tanjungpinang dapat diketahui dengan cara mempelajari pola cuaca menggunakan data data yang telah lalu untuk mendapatkan klasifikasi keadaan cuaca. Klasifikasi keadaan cuaca dapat menggunakan algoritma *K-Nearest Neighbor*. *K-Nearest Neighbor* merupakan suatu algoritma *supervised learning* dimana hasil yang didapat diklasifikasikan berdasarkan mayoritas dari jarak tetangga yang terdekat (Maulana, 2018).

Berdasarkan penjelasan di atas, penulis ingin melakukan penelitian skripsi yang berjudul “Penerapan Metode *K-Nearest Neighbor* (KNN) Untuk Klasifikasi Cuaca Di Wilayah Tanjungpinang”. Pada penelitian ini, pembuatan sistem digunakan untuk dapat mengetahui pola cuaca di Tanjungpinang.

II. Metode Penelitian

2.1 Normalisasi

Menurut Hanifa, *et al.* (2017) Normalisasi merupakan suatu proses yang dilakukan untuk meningkatkan kinerja dalam model klasifikasi serta memudahkan dalam proses klasifikasi. Data atribut diskalakan dalam *range* 0 sampai 1. Normalisasi terdapat banyak jenis, salah satunya adalah *Normalization min-max method*. Pada normalisasi *min-max* adalah metode kompleks yang tidak menghilangkan isi dari data yang dinormalisasi.

Menurut Wiyono dan Abidin (2018) berikut merupakan rumus normalisasi data :

$$X_{new} = \frac{x - \min(x)}{\max(x) - \min(x)}$$

Keterangan:

X_{new} = data baru

$\min(x)$ = data dengan nilai terendah

$\max(x)$ = data dengan nilai tertinggi

2.2 K-Nearest Neighbor

Algoritma *K-Nearest Neighbor* merupakan algoritma *supervised learning* dimana hasil yang didapat diklasifikasikan berdasarkan mayoritas dari jarak tetangga yang terdekat. Tujuan algoritma ini untuk mengklasifikasikan objek baru didapat berdasarkan atribut dan sampel dari data pelatihan. Pada umumnya nilai k menggunakan jumlah ganjil agar tidak terdapat jarak yang sama dalam proses klasifikasi. Jauh atau dekatnya tetangga dihitung menggunakan jarak *Euclidean Distance* (Maulana, 2018) .

Menurut Nasution, *et al.* (2019) berikut merupakan rumus dari *Euclidean Distance*:

$$d(x,y) = \sqrt{(\sum_{i=1}^n (x_i - y_i)^2)}$$

Keterangan:

X_i = data baru

Y_i = data pelatihan

i = variabel data

$d(x, y)$ = jarak antara x dengan y

n = dimensi data

2.3 Data Mining

Data *mining* adalah proses yang memiliki tahap tahap interaktif yang berhubungan dengan *knowledge base* (Elisa, 2017). Data *mining* merupakan proses yang digunakan untuk dapat menemukan pola atau hubungan yang terjadi dalam data.

Menurut Banjarsari, *et al.* (2015) Didalam data mining terdapat beberapa proses yaitu :

1. pencarian k-Optimal pada algoritma KNN menggunakan *Fold Cross Validation*.
2. Setelah mendapat nilai k-Optimal nilai tersebut akan digunakan untuk diuji akurasi menggunakan data real.
3. Melakukan prediksi dengan menggunakan algoritma KNN menggunakan nilai k hasil dari *k-Folds Cross Validation* dengan variabel *input*.

Berikut merupakan rumus yang digunakan untuk melakukan validasi:

$$Accuracy = \frac{\text{jumlah nilai benar}}{\text{jumlah data keseluruhan}} \times 100\%$$

2.4 Klasifikasi

Menurut Hutami (2016) menjelaskan bahwa klasifikasi merupakan teknik untuk memetakan kelas kedalam data yang telah terklasifikasi dan terdapat 3 tahapan dalam melakukan klasifikasi yaitu :

1. Perancangan model

Merupakan langkah awal yang dilakukan sebelum melakukan permodelan. Perancangan digunakan agar hasil yang diinginkan dapat di rancang atau diduga sebelum menjadi hasil yang baik.

2. Implementasi model

Implementasi model adalah tahapan yang dilakukan untuk menerapkan dari perancangan model. Implementasi merupakan proses yang dilakukan untuk membentuk hasil setelah direncanakan.

3. Evaluasi model

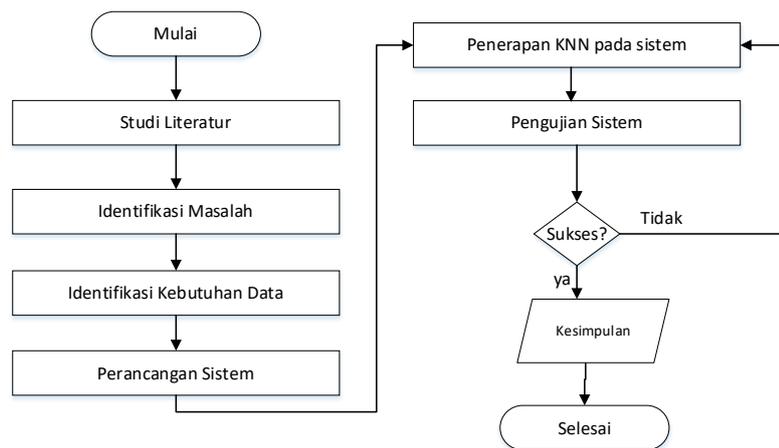
Evaluasi model adalah tahapan yang dilakukan untuk memperbaiki atau menambah fitur yang telah dibuat. Evaluasi dapat digunakan untuk dilakukan pengembangan yang lebih baik lagi.

2.5 Cuaca

Cuaca adalah suatu kondisi yang menggambarkan atau mewakili kondisi udara yang ada di bumi. Cuaca dapat berubah ubah dari waktu ke waktu. Cuaca merupakan (ilmu yang mempelajari cuaca yaitu meteorologi) yang dapat diartikan dengan memahami kondisi fisik pada suatu daerah atau lokasi dan dalam waktu tertentu. Susunan dalam unsur cuaca meliputi : intensitas radiasi matahari, lama penyinaran matahari, suhu, udara, kelembapan udara, tekanan udara, kecepatan angin, arah angin, penutupan awan, *presipitasi* dan *evapotranspirasi* (Kaho, 2014)

2.6 Flowchart Prosedur Penelitian

Adapun Flowchart Prosedur Penelitian yang dilakukan pada penerapan metode *K-Nearest Neighbor* seperti terlihat pada Gambar 1.



Gambar 1. Flowchart Prosedur Penelitian

a. Studi Literatur

Pada tahap studi literatur peneliti mencari sumber bacaan dan referensi dari berbagai sumber tertulis seperti artikel, buku, jurnal atau dokumen yang relevan berkaitan dengan penelitian.

b. Identifikasi Masalah

Pada tahap identifikasi masalah peneliti melakukan pengenalan terhadap masalah yaitu bagaimana menerapkan algoritma K-Nearest Neighbor (KNN) untuk klasifikasi cuaca di wilayah Tanjungpinang.

c. Identifikasi Kebutuhan Data

Pada tahap Identifikasi kebutuhan data peneliti mencari data yang diperlukan di Stasiun Meteorologi Raja Haji Fisabilillah. Data yang dicari antara lain memiliki parameter suhu udara, kelembapan udara, tekanan udara dan kecepatan angin.

d. Penerapan KNN Pada Sistem

Pada tahap penerapan KNN pada sistem peneliti menerapkan metode KNN kedalam sistem.

- e. Perancangan Sistem
Pada tahap perancangan sistem peneliti membuat suatu rancangan sistem yaitu meliputi perancangan *Diagram konteks, Data Flow Diagram, Entity Relationship Diagram*.
- f. Penerapan KNN Pada Sistem
Pada tahap penerapan KNN pada sistem peneliti menerapkan metode KNN kedalam sistem.
- g. Pengujian Sistem
Pada tahap ini peneliti melakukan uji metode untuk mengetahui apakah sistem yang dibangun mengalami kegagalan atau *error* juga dilakukan uji untuk mengetahui hasil klasifikasi serta untuk mengetahui tingkat akurasi.
- h. Kesimpulan
Pada tahap kesimpulan peneliti menyimpulkan hasil dari penelitian penerapan metode K-Nearest Neighbor (KNN) untuk klasifikasi cuaca di wilayah Tanjungpinang.

2.5 Flowchart K-Nearest Neighbor

Adapun Flowchart Prosedur Penelitian yang dilakukan pada implementasi Metode DBSCAN seperti terlihat pada Gambar 2.



Gambar 2. Flowchart KNN

- a. Masukkan Data Sampel
Bagian masukkan data sampel merupakan proses memasukkan data cuaca wilayah Tanjungpinang pada periode Januari 2016 hingga Desember 2019 yang diperoleh dari Stasiun Meteorologi Raja Haji Fisabilillah.
- b. Masukkan Data Training
Bagian masukkan data training merupakan proses memasukkan data latih atau *training* sebagai pembelajaran yang digunakan untuk menentukan klasifikasi objek baru pada data *testing* atau data uji.

- c. Menghitung Jarak
Bagian menghitung jarak merupakan proses untuk menghitung jarak antara data *training* dengan data *testing*.
- d. Urut Data Hasil Perhitungan Mulai dari Jarak Terkecil
Bagian mengurutkan data hasil perhitungan mulai dari jarak terkecil merupakan proses untuk mengurutkan hasil perhitungan mulai dari yang terkecil.
- e. Menentukan Jumlah K
Bagian menentukan jumlah k merupakan proses menentukan jumlah tetangga terdekat. K merupakan nilai parameter tetangga terdekat dengan data *testing* yang akan *ditraining*. Contohnya misal k=13 maka hasil dari klasifikasi masuk kekelas mayoritas 13 tetangga yang terdekat.
- f. Mencari Mayoritas Kelas Data Testing Berdasarkan dengan jumlah K
Merupakan proses untuk mencari objek pada data testing yang baru diuji berdasarkan mayoritas kelas yang muncul sesuai dengan k yang telah ditentukan.
- g. Hasil Klasifikasi Cuaca
Hasil klasifikasi ditampilkan dari data testing yang sudah diuji mulai dari hasil yang terkecil.

III. Hasil dan Pembahasan

3.1 Hasil Penelitian

Data yang digunakan merupakan data pada Stasiun Meteorologi Raja Haji Fisabilillah Tanjungpinang mulai dari Januari tahun 2016-Desember tahun 2019. Jumlah data yang digunakan sebanyak 1450 data.

3.2 Pengujian dan Analisa

Proses pengujian akan dilakukan dengan menggunakan 10 data cuaca di wilayah Tanjungpinang.

Tabel 1. Hasil Pengujian dengan menggunakan k=13

id	Training	Knn	validasi
1452	5	7	Tidak sama
1453	1	7	Tidak sama
1454	5	7	Tidak sama
1455	5	7	Tidak sama
1456	5	5	Sama
1457	5	5	Sama
1458	7	5	Tidak sama
1459	5	5	Sama
1460	5	5	Sama
1461	5	5	Sama

Maka perhitungan persentase kesamaan data untuk nilai K=13 adalah sebagai berikut:

$$\begin{aligned} \text{Persentase Akurasi (K=13)} &= \frac{5}{10} \times 100\% \\ &= 50\% \end{aligned}$$

Dari 1450 data training dan 10 data testing, didapatkan hasil perbandingan kesamaan hasil klasifikasi cuaca oleh Stasiun Meteorologi Raja Haji Fisabilillah Tanjungpinang dengan algoritma KNN yang bernilai K=13 memiliki kesamaan persentase yaitu sebesar 50%.

3.3 Pembahasan

Proses pengujian dilakukan dengan menggunakan 10 data dan menggunakan beberapa nilai k.

Tabel 2. Hasil Pengujian dengan menggunakan beberapa k.

No	Nilai K	Persentase Kesamaan/Akurasi (%)
1.	5	40%
2.	7	30%
3.	9	60%
4.	11	60%
5.	13	50%
6.	15	50%
7.	17	50%
8.	19	60%
9.	21	60%
10.	23	60%
11.	25	60%
12.	27	60%
13.	29	60%
14.	31	70%
15.	33	70%
16.	35	70%
17.	37	70%
18.	39	70%
19.	41	70%
20.	43	60%
21.	45	70%
22.	47	70%
23.	50	70%

Tabel 2 merupakan persentase kesamaan hasil klasifikasi yang dilakukan dengan menggunakan perhitungan manual metode *K-Nearest Neighbor* (KNN), maka didapatkan persentase tertinggi yaitu dengan nilai K= 31, 33, 35, 37, 39, 41, 45, 47, 50 tingkat persentase kesamaan sebesar 70%.

IV. Kesimpulan

Kesimpulan yang dapat diambil dari penelitian ini adalah :

1. Penerapan algoritma *K-Nearest Neighbor* untuk klasifikasi cuaca di wilayah Tanjungpinang telah berhasil diterapkan dengan melakukan langkah-langkah seperti yang sudah dilakukan dan menghasilkan akurasi tertinggi sebesar 70% dan akurasi terendah sebesar 30% sehingga hanya memiliki selisih sebesar 40%.
2. Berdasarkan dari pengujian yang dilakukan dapat diketahui bahwa *K-Nearest Neighbor* sudah memberikan nilai yang baik.

V. Daftar Pustaka

- Banjarsari, M.A., Budiman, I., dan Farmadi, 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 (KLIK)*, Banjarnegara.
- Elisa, E., 2017, Analisa dan Penerapan Algoritma C4.5 dalam Data Mining untuk Mengidentifikasi faktor-faktor penyebab kecelakaan kerja Kontruksi PT.Arupadhatu Adisesanti , *Jurnal Online Informatika*, 2. 36. 10.15575.
- Hanifa, Tesha, T., dan Adiwijaya., dan Faraby, Said, A., 2017, Analisis *Churn Prediction* pada Data Pelanggan PT.Telekomunikasi dengan Logistic *Regression* dan *Underbagging*, *Jurnal e-Proceeding of Engineering*.
- Hutami, R., dan Astuti, E.Z., 2016, Implementasi Metode K-Nearest Neighbor Penjualan Furniture

- Pada Cv. Octo Agung Jepara, *Dokumen Karya Ilmiah*, Semarang.
- Kaho, Norman, P., 2014. Panduan Interpretasi dan Respon Informasi Iklim dan Cuaca untuk Petani dan Nelayan. Kupang: Penerbit:Perkumpulan Pikul
- Maulana, Asep., 2018. cara kerja Algoritma k-Nearest Neighbor (k-NN). <https://medium.com/bee-solution-partners/cara-kerja-algoritma-k-nearest-neighbor-k-nn-389297de543e>. Diakses pada 28 Februari 2020 jam 13.25 WIB.
- Nurhadi., Saleh, B.A., dan Badri, D.A., dan Susanti, P., 2009, *Jelajah Cakrawala Sosial*, Jakarta: CV. Citra Praya
- Wiyono, S., dan Abidin, T., 2018, *Implementation of K-Nearest Neighbor (KNN) Algorithm to Predict Student's Performance*, Jurnal SIMETRIS, Tegal.