

**METODE *k*-NEAREST NEIGHBOR (*k*-NN) REGRESSION DALAM
MEMPREDIKSI INDEKS PEMBANGUNAN MANUSIA**

(Skripsi)

Oleh

Levia Febriyanti



**JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2022**

ABSTRACT

THE *k*-NEAREST NEIGHBOR (*k*-NN) REGRESSION METHOD IS USED TO PREDICT THE HUMAN DEVELOPMENT INDEX

By

Levia Febriyanti

The *k*-Nearest Neighbor (*k*-NN) method is one of the *Machine Learning* (ML) algorithms that can be used for classification and regression analysis. The *k*-NN method is considered a simple method to be applied in data analysis with multiple variable dimensions. In this research, the *k*-NN *regression* method was chosen with the aim of predicting the best *k* value and *Mean Absolute Percentage Error* (MAPE) on Human Development Index (HDI) in Indonesia for 2022. The data used in this research consisted of HDI, life expectancy index, mean years of schooling, expected years of schooling, and gross national index per capita from 2020-2021 of 34 provinces in Indonesia. The results of this research indicate that the optimal value of *k* is $k = 2$ with an MAPE error of 0,9019%. The average HDI prediction in Indonesia for 2022 is 71,41.

Key Words: Regression analysis, multiple variable, *k*-Nearest Neighbor (*k*-NN) *Regression*, *Mean Absolute Percentage Error* (MAPE).

ABSTRAK

METODE *k*-NEAREST NEIGHBOR (*k*-NN) REGRESSION DALAM MEMPREDIKSI INDEKS PEMBANGUNAN MANUSIA

Oleh

Levia Febriyanti

Metode *k*-Nearest Neighbor (*k*-NN) merupakan salah satu algoritma *Machine Learning* (ML) yang dapat digunakan untuk klasifikasi dan analisis regresi. Metode *k*-NN dianggap sebagai metode yang sederhana untuk diterapkan dalam analisis data dengan dimensi peubah banyak. Pada penelitian ini metode *k*-NN *Regression* dipilih dengan tujuan untuk memprediksi nilai *k* dan *Mean Absolute Percentage Error* (MAPE) terbaik pada Indeks Pembangunan Manusia (IPM) di Indonesia tahun 2022. Data yang digunakan pada penelitian ini terdiri dari data IPM, angka harapan hidup, rata-rata lama sekolah, harapan lama sekolah, dan pengeluaran perkapita riil pertahun yang disesuaikan dari tahun 2020-2021 sebanyak 34 provinsi di Indonesia. Hasil penelitian ini menunjukkan bahwa nilai *k* yang optimal adalah $k = 2$ dengan galat MAPE 0,9019% diperoleh hasil prediksi IPM di Indonesia pada tahun 2022 rata-rata sebesar 71,41.

Kata Kunci: Analisis Regresi, Peubah Banyak, *k*-Nearest Neighbor (*k*-NN) *Regression*, *Mean Absolute Percentage Error* (MAPE).

**METODE *k*-NEAREST NEIGHBOR (*k*-NN) REGRESSION DALAM
MEMPREDIKSI INDEKS PEMBANGUNAN MANUSIA**

Oleh

LEVIA FEBRIYANTI

Skripsi

Sebagai Salah Satu Syarat untuk Mencapai Gelar
SARJANA MATEMATIKA

Pada

Jurusan Matematika
Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Lampung



**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
2022**

Judul Skripsi : **METODE *k*-NEAREST NEIGHBOR (*k*-NN) REGRESSION DALAM MEMPREDIKS INDEKS PEMBANGUNAN MANUSIA**

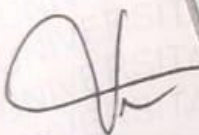
Nama Mahasiswa : **Levia Febriyanti**

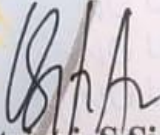
Nomor Pokok Mahasiswa : **1617031089**

Jurusan : **Matematika**

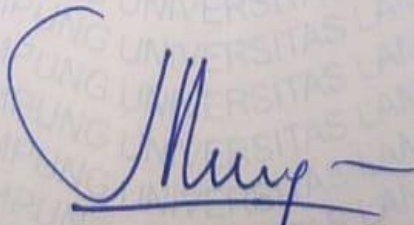
Fakultas : **Matematika dan Ilmu Pengetahuan Alam**




Drs. Nusyirwan, S.Si., M.Si.
NIP. 19661010 199205 1 001


Dr. Fitriani, S.Si., M.Sc.
NIP. 19840627 200604 2 001

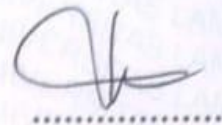
2. Mengetahui
Ketua Jurusan Matematika


Dr. Aang Nuryaman, S.Si., M.Si.
NIP. 19740316 200501 1 001

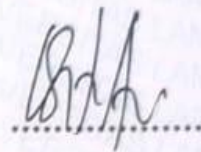
MENGESAHKAN

1. Tim Penguji


Ketua : Drs. Nusyirwan, S.Si., M.Si.



Sekretaris : Dr. Fitriani, S.Si., M.Sc.



Penguji
Bukan Pembimbing : Dr. Khoirin Nisa, S.Si., M.Si.



2. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam




Dr. Eng Satripto Dwi Yuwono, S.Si., M.T.
NIP. 19740705 200002 1 001

Tanggal Lulus Ujian Skripsi : 25 Juli 2022

PERNYATAAN SKRIPSI MAHASISWA

Nama : **Levia Febriyanti**
Nomor Pokok Mahasiswa : **1617031089**
Jurusan : **Matematika**
Judul : **Metode *k-Nearest Neighbor (k-NN) Regression*
Dalam Memprediksi Indeks Pembangunan
Manusia**

Dengan ini saya menyatakan bahwa skripsi ini adalah hasil pekerjaan saya sendiri dan menurut pengetahuan saya juga tidak terdapat karya atau pendapat yang ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini sebagaimana disebutkan dalam daftar pustaka. Serta semua tulisan yang tertuang dalam skripsi ini telah mengikuti kaidah karya penulisan ilmiah Universitas Lampung.

Bandar Lampung, 25 Juli 2022



Levia Febriyanti
NPM. 1617031089

RIWAYAT HIDUP

Penulis dilahirkan pada tanggal 23 Februari 1998 di Serang, sebagai anak kedua dari tiga bersaudara pasangan Bapak Samud dan Ibu Sri Yulianingsih, S.Pd., adik dari Rafi'i Zaidan Rifani, S.T. dan kakak dari Fitra Nurindra.

Penulis telah menyelesaikan pendidikan TK di TK Artha Kencana pada tahun 2004, pendidikan sekolah dasar di SDN 03 Serang pada tahun 2010, pendidikan menengah pertama di SMP La Tansa pada tahun 2013, pendidikan menengah atas di SMAN 2 Kota Serang pada tahun 2016 dan pada tahun 2016 penulis diterima sebagai mahasiswi di Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung melalui jalur Seleksi Bersama Masuk Perguruan Tinggi Negeri (SBMPTN).

Selama menjadi mahasiswa, penulis aktif di Himpunan Mahasiswa Jurusan Matematika (HIMATIKA) sebagai Anggota Bidang Kaderisasi dan Kepemimpinan periode 2017 dan Staff Ahli Kementerian Luar Negeri Kabinet Sinergis dalam Gerak di BEM U KBM Unila periode 2018.

Sebagai bentuk aplikasi bidang ilmu kepada masyarakat, penulis telah menyelesaikan Kerja Praktik (KP) selama satu bulan di Badan Pusat Statistik Kota Serang, serta Kuliah Kerja Nyata (KKN) selama 40 hari di Desa Subik, Kecamatan Abung Tengah, Kabupaten Lampung Utara.

Dengan mengucapkan Alhamdulillah Rabbil 'Alamiin ,
Puji dan syukur kehadiran Allah Subhanahu Wata'ala segala nikmat dan karunia-
Nya, dan suri tauladan Nabi Muhammad Shallallahu 'Alaihi Wassallam yang
menjadi contoh dan panutan untuk kita semua.

Kupersembahkan sebuah karya sederhana ini untuk:

Ayahanda Samud dan Ibunda Sri Yulianingsih, S.Pd.

Terimakasih atas limpahan kasih sayang, pengetahuan, doa dan seluruh motivasi
disetiap langkahku. Karena atas doa dan ridho kalian, Allah memudahkan setiap
perjalanan hidup ini.

Terimalah bukti kecil ini sebagai hadiah keseriusanku untuk membalas semua
pengorbanan, keikhlasan, dan jerih payah yang selama ini kau lakukan.

Kakak dan Adik Tercinta

Rafi'i Zaidan Rifani, S.T.

Fitra Nurindra

Almamaterku Tercinta

Universitas Lampung

KATA INSPIRASI

*Ku biarkan gelombang
Tumbuh di kepalaku
Badai akan ia hadang
Jurang tak akan bikin mudur*

*Biarkan gelombang
Memanjang di kepalaku
Liar hingga palung dalam
Juang meski nyaris tenggelam*

*Tak ada mesin penghenti aku
Besi-besi karatan batu mengapur
Tapi aku tak akan sejangkal pudur*

(Boy Candra)

SANWACANA

Alhamdulillah Rabbil ‘Alamiin, puji syukur penulis panjatkan atas kehadiran Allah SWT. yang telah melimpahkan rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi yang judul “Metode *k-Nearest Neighbor (k-NN) Regression* Dalam Memprediksi Indeks Pembangunan Manusia”.

Terselesaikannya skripsi ini tidak terlepas dari bantuan, kerjasama, dan dukungan dari berbagai pihak. Untuk itu penulis mengucapkan terimakasih kepada:

1. Bapak Drs. Nusyirwan, M.Si., selaku dosen pembimbing utama yang telah meluangkan waktu dari padatnya kesibukan beliau, dalam membimbing dan memotivasi penulis selama melaksanakan penelitian dan penyelesaian skripsi.
2. Ibu Dr. Fitriani, S.Si., M.Sc., selaku dosen pembimbing kedua yang telah banyak membantu dan memberikan pengarahan dalam proses penyusunan skripsi.
3. Bapak Drs. Rudi Ruswandi, M.Si. dan Ibu Dr. Khorin Nisa, S.Si., M.Si., selaku dosen penguji yang telah memberikan nasihat, motivasi, saran dan kritik yang membangun guna penyempurnaan skripsi ini.
4. Bapak Dr. Muslim Ansori, M.Si., selaku dosen pembimbing akademik yang telah membimbing penulis selama menjalankan studi di jurusan matematika.

5. Bapak Drs. Aang Nuryaman, M.Si., selaku ketua Jurusan matematika FMIPA Universitas Lampung.
6. Bapak Dr. Eng. Suropto Dwi Yuwono, S.Si., M.T., selaku dekan FMIPA Universitas Lampung.
7. Seluruh dosen, staff, dan karyawan Jurusan Matematika FMIPA Universitas Lampung yang telah memberikan banyak ilmu pengetahuan dan bantuan kepada penulis.
8. Papah, mamah, a Rivan, dek Indra, dan keluarga yang telah memberikan motivasi dan bantuan baik moril maupun materil dan memberikan segala perhatiannya serta selalu mendoakan agar penulis dapat menyelesaikannya.
9. Sahabat-sahabat penulis Alif, Jeje, Shofi, Salsa, Tiyas yang senantiasa menemani suka duka penulis dan selalu menyemangati penulis.
10. Keluarga WHS Nuurul, Stacia, Azkia, Ayu, Dhara, Dian, Elly, Mei, Ami, Anggun, Intan, Irfan, Desfan, Desi, Laras, Sam dan teman-teman angkatan 2016, adik-adik, serta abang dan yunda yang telah menemani disetiap perjalanan penulis dari mahasiswa baru hingga sarjana.
11. Semua pihak yang yang telah membantu selama ini, yang tidak dapat disebutkan satu persatu.

Bandar Lampung, 25 Juli 2022
Penulis

Levia Febriyanti

DAFTAR ISI

	Halaman
DAFTAR TABEL	iii
DAFTAR GAMBAR	iv
I. PENDAHULUAN	
1.1 Latar Belakang dan Masalah	1
1.2 Tujuan Penelitian	4
1.3 Manfaat Penelitian	4
II. TINJAUAN PUSTAKA	
2.1 Peramalan	5
2.2 Jarak <i>Euclidean</i>	6
2.3 Data Latih dan Data Uji	7
2.4 Pola Data	8
2.5 Metode <i>k-Nearest Neighbor (k-NN) Regression</i>	11
2.6 <i>Mean Absolute Percentage Error (MAPE)</i>	15
2.7 Indeks Pembangunan Manusia	16
2.7.1 Indeks Kesehatan	17
2.7.2 Indeks Pendidikan	18
2.7.3 Indeks Standar Hidup Layak	19
III. METODOLOGI PENELITIAN	
3.1 Waktu dan Tempat Penelitian	21
3.2 Data Penelitian	21
3.3 Metode Penelitian	22
IV. HASIL DAN PEMBAHASAN	
4.1 Pemilihan Data	23
4.2 Perhitungan Metode <i>k-NN Regression</i>	30
4.2.1 Menghitung Jarak <i>Euclidean</i>	30
4.2.2 Mengurutkan Jarak	31
4.2.3 Menentukan Nilai <i>k</i>	33
4.2.4 Menghitung Rata-Rata (\hat{Y}_i)	34
4.3 Pemilihan Prediksi Terbaik	38
4.4 Analisis Hasil Prediksi Terbaik	40

V. KESIMPULAN
DAFTAR PUSTAKA
LAMPIRAN

DAFTAR TABEL

Tabel	Halaman
1. Urutan jarak <i>Euclidean</i> data latih	31
2. Urutan jarak <i>Euclidean</i> data uji	32
3. Nilai MAPE hasil prediksi IPM di Indonesia	38
4. Prediksi Nilai IPM di Indonesia tahun 2020	40

DAFTAR GAMBAR

Gambar	Halaman
1. Pola data horisontal.....	9
2. Pola data musiman	9
3. Pola data siklus	10
4. Pola data <i>trend</i>	10
5. <i>K</i> -NN dengan nilai: (a) $K = 1$, (b) $K = 2$, (c) $K = 5$, (d) $K = 7$	14
6. Grafik indeks pembangunan manusia di Indonesia	23
7. Grafik angka harapan hidup di Indonesia	25
8. Grafik rata-rata lama sekolah di Indonesia.....	26
9. Grafik harapan lama sekolah di Indonesia	27
10. Grafik pendapatan per-kapita riil pertahun di Indonesia.....	28
11. Hasil prediksi IPM di Indonesia tahun 2020	41

I. PENDAHULUAN

1.1 Latar Belakang dan Masalah

Pembangunan di suatu daerah merupakan suatu upaya yang dilakukan oleh pemerintah daerah terkait kesejahteraan masyarakat serta mewujudkan kemakmuran masyarakat. Salah satu pembangunan yang menjadi perhatian pemerintah daerah adalah pembangunan manusia yang diukur menggunakan indikator yang disebut Indeks Pembangunan Manusia (Putra, 2015).

Indeks Pembangunan Manusia (IPM) merupakan tingkatan status pembangunan manusia di suatu wilayah yang akan berfungsi sebagai patokan dasar perencanaan jika dibandingkan dengan antar waktu untuk memberikan gambaran kemajuan setelah suatu periode atau antar wilayah untuk memberikan gambaran tentang tingkat kemajuan suatu wilayah relatif terhadap wilayah lain. Ada tiga dimensi dasar penyusun IPM yang meliputi indeks kesehatan, indeks pendidikan dan indeks standar hidup layak (BPS, 2008). Indeks kesehatan meliputi angka harapan hidup, indeks pendidikan meliputi indikator harapan lama sekolah dan rata-rata lama sekolah, dan indeks standar hidup layak meliputi indikator kemampuan daya beli masyarakat terhadap sejumlah kebutuhan pokok yang dilihat dari rata-rata besarnya pengeluaran per kapita yang disesuaikan.

Karena begitu pentingnya IPM di Indonesia, sehingga perlu dilakukan prediksi tingkat perkembangan IPM di tahun-tahun selanjutnya agar pemerintah Indonesia memiliki referensi dan acuan yang jelas untuk menentukan kebijakan ataupun membuat langkah-langkah strategis yang tepat agar IPM di Indonesia tidak menurun di masa yang akan datang, bahkan meningkat pada tiap tahunnya.

Berdasarkan data Badan Pusat Statistik, secara garis besar IPM tahun 2019 hingga tahun 2020 mengalami peningkatan, namun tidak terlalu signifikan. Seperti pada tahun 2019 tingkat IPM di Indonesia sebesar 71,92, sedangkan pada tahun 2020 sebesar 71,94 atau meningkat sebesar 2%. Namun pada provinsi-provinsi di Indonesia tingkat IPM sangat beragam seperti halnya pada tahun 2020 di Provinsi Papua tingkat IPM sebesar 60,44 dan tingkat IPM di Provinsi DKI Jakarta sebesar 80,77. Oleh karena itu, dibutuhkan metode-metode yang sudah teruji agar keakuratan yang dihasilkan dapat dipertanggungjawabkan. Guna mempermudah pemerintah Indonesia dalam memprediksi tingkat IPM provinsi-provinsi di Indonesia, dibutuhkan metode *k-Nearest Neighbor (k-NN) Regression* yang tepat. Upaya pemerintah Indonesia dapat menanggulangi ketidakseragaman tingkat IPM di 34 provinsi Indonesia harus dilakukan secara adil menyeluruh agar seluruh masyarakat Indonesia hidupnya sejahtera.

Metode *k-NN* merupakan salah satu algoritma *Machine Learning (ML)* yang dapat digunakan untuk memprediksi nilai IPM. Metode *k-NN* dianggap sebagai metode yang sederhana untuk diterapkan dalam analisis data dengan dimensi peubah banyak (Alkhatib, 2013).

Awalnya k -NN merupakan metode untuk analisis klasifikasi, namun sering juga digunakan untuk regresi. Walaupun metode k -NN dianggap sederhana namun metode ini memiliki kelebihan, yaitu dapat menggeneralisasi himpunan data latih yang relatif kecil (Rokach, 2010).

Seperti pada penelitian terdahulu yang dilakukan oleh Sinta (2014) yang memprediksi harga beras di Indonesia dengan menggunakan metode *Ensemble k-NN* menghasilkan hasil prediksi harga yang hampir sama dengan harga aktualnya. Penelitian berikutnya yaitu penelitian oleh Fauzi (2017), pada penelitian tersebut membahas pengklasifikasian atau pengelompokkan IPM dengan variabel harapan lama sekolah sebagai faktor paling besar mempengaruhi IPM dengan menggunakan metode k -NN dan metode SVM (*Support Vector Machine*). Selain itu penelitian lainnya tentang regresi k -NN yang dilakukan oleh Seruni dkk. (2020) dalam memprediksi jumlah penduduk di kota Malang. Berdasarkan hasil penelitian terkait, dapat disimpulkan bahwa metode k -NN memiliki hasil yang baik dalam prediksi nilai yang dibuktikan dengan nilai *error* yang kecil.

Berdasarkan latar belakang yang telah diuraikan sebelumnya, penulis tertarik untuk melakukan penelitian dengan judul “Metode *k-Nearest Neighbor (k-NN) Regression* dalam Memprediksi Indeks Pembangunan Manusia”. Penulis mengambil studi kasus di Indonesia dan data yang digunakan diambil dari Badan Pusat Statistik (BPS).

1.2 Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah:

1. memprediksi nilai k terbaik Indeks Pembangunan Manusia di Indonesia pada Tahun 2022 menggunakan metode *k-NN Regression*,
2. mengetahui nilai *Mean Absolute Percentage Error* (MAPE) pada Indeks Pembangunan Manusia di Indonesia tahun 2022 menggunakan metode *k-NN Regression*.

1.3 Manfaat Penelitian

Manfaat dari penelitian ini adalah:

1. mengetahui cara kerja metode *k-NN Regression* dalam memprediksi suatu data,
2. menambah referensi dalam pengembangan atau penelitian selanjutnya yang membahas tentang prediksi dan metode *k-NN Regression*.

II. TINJAUAN PUSTAKA

2.1 Peramalan

Menurut Subagyo (1986), peramalan adalah perkiraan mengenai sesuatu yang belum terjadi. Dalam ilmu pengetahuan sosial segala sesuatu itu serba tidak pasti, harus diperkirakan dengan tepat. Dalam hal ini perlu dilakukan peramalan berdasarkan data yang terdapat di masa lalu yang dianalisis menggunakan cara-cara tertentu. Peramalan menjadi bagian penting dalam setiap organisasi, hal ini terkait setiap pengambilan keputusan manajemen yang tentunya akan mempengaruhi perkembangan organisasi. Peramalan menjadi dasar bagi perencanaan jangka panjang maupun jangka pendek dari suatu organisasi (Arga, 1984).

Metode peramalan dibagi ke dalam dua kategori utama, yaitu metode kuantitatif dan metode kualitatif. Metode kuantitatif dilakukan apabila informasi masa lalu tersedia sehingga peramalan bisa dilakukan, informasi tersebut dapat dikuantitatifkan dalam bentuk data numerik.

Dalam metode kualitatif pendapat-pendapat dari para ahli akan menjadi pertimbangan dalam pengambilan keputusan sebagai hasil dari peramalan yang telah dilakukan. Apabila ada data masa lalu tersedia, peramalan dengan metode kuantitatif akan lebih efektif digunakan dibandingkan dengan metode kualitatif (Makridakis dkk., 1999).

2.2 Jarak *Euclidean*

Jarak *Euclidean* adalah metrika yang paling sering digunakan untuk menghitung kesamaan dua vektor. Jarak *Euclidean* merupakan perhitungan akar dari kuadrat perbedaan dua vektor (Taufiqurrohman, 2013). Fungsi jarak yang paling umum adalah jarak *Euclidean*, yang mewakili cara biasa manusia berpikir tentang jarak di dunia nyata. Jarak *Euclidean* merupakan jarak geometris antar dua objek data. Jika semakin dekat jarak, maka semakin mirip suatu objek data tersebut. Jarak *Euclidean* dapat dihitung sebagai berikut (Kataria dan Singh, 2013).

Jika dua vektor x_i dan x_j diberikan, dimana:

$$x_i = \{x_{i1}, x_{i2}, x_{i3}, x_{i4}, \dots, x_{ip}\}$$

$$x_j = \{x_{j1}, x_{j2}, x_{j3}, x_{j4}, \dots, x_{jp}\}$$

sehingga jarak antara x_i dan x_j

$$D_{(i,j)} = \sqrt{(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 + \dots + (x_{ip} - x_{jp})^2} \quad (2.1)$$

Persamaan 2.1 dapat dibentuk juga menjadi persamaan sebagai berikut:

$$D_{(i,j)} = \sqrt{\sum_{l=1}^p (x_{il} - x_{jl})^2} \quad (2.2)$$

dengan:

$D_{i,j}$ = jarak *Euclidean*,

x_i = nilai variabel objek ke- i ,

x_{ip} = nilai variabel objek ke- i atribut ke- p ,

p = banyaknya atribut,

n = banyaknya data yang diamati,

i, j = $1, 2, \dots, n$,

l = $1, 2, \dots, p$.

Menurut Santosa (2007), ukuran jarak harus memenuhi syarat-syarat sebagai berikut:

1. $D(x_i, x_j) \geq 0$ (tidak negatif), tidak ada jarak yang mempunyai nilai negatif,
2. $D(x_i, x_j) = 0$ jika dan hanya jika $x_i = x_j$, jarak antara suatu titik dengan titik itu sendiri adalah 0,
3. $D(x_i, x_j) = D(x_j, x_i)$, jarak dari x_i ke x_j adalah sama dengan jarak x_j ke x_i .

2.3 Data Latih dan Data Uji

Data latih (x^{latih}) adalah kumpulan data yang digunakan selama proses pembelajaran dan digunakan agar sesuai dengan bobot dari sebuah pengelompokkan.

Sebagian besar pendekatan yang mencari melalui data pelatihan untuk hubungan empiris cenderung menghasilkan yang terbaik, yang berarti bahwa mereka dapat

mengidentifikasi dan mengeksploitasi hubungan nyata dalam data pelatihan yang tidak berlaku secara umum.

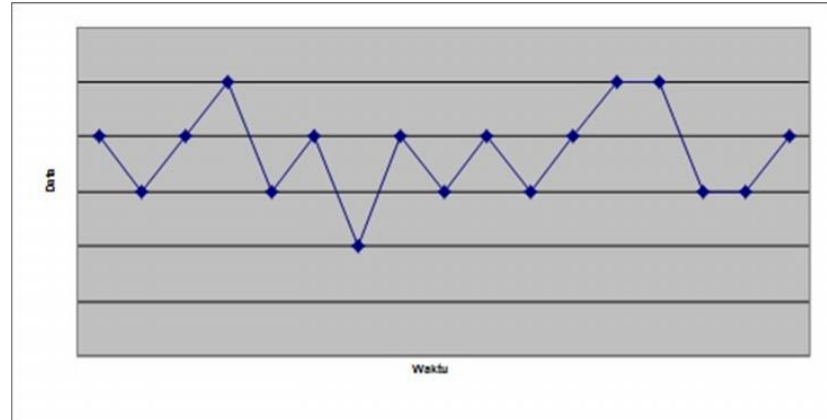
Data uji (x^{uji}) adalah dataset yang tidak tergantung pada dataset pelatihan. Jika model cocok dengan dataset pelatihan, maka model juga cocok dengan dataset uji dengan baik, kecocokkan model minimal telah terjadi. Pemasangan yang lebih baik dari dataset pelatihan dibandingkan dengan dataset uji biasanya menunjuk hasil yang terbaik. Oleh karena itu, data uji adalah serangkaian data yang hanya digunakan untuk menilai kinerja dari pengelompokan yang ditentukan sepenuhnya (Ripley, 1996).

2.4 Pola Data

Menurut Makridakis dkk. (1999), hal penting yang harus diperhatikan dalam metode deret waktu adalah menentukan jenis pola data, sehingga dapat dengan tepat menentukan metode yang akan digunakan. Pola data dapat dibedakan sebagai berikut.

1. Pola Data Horisontal

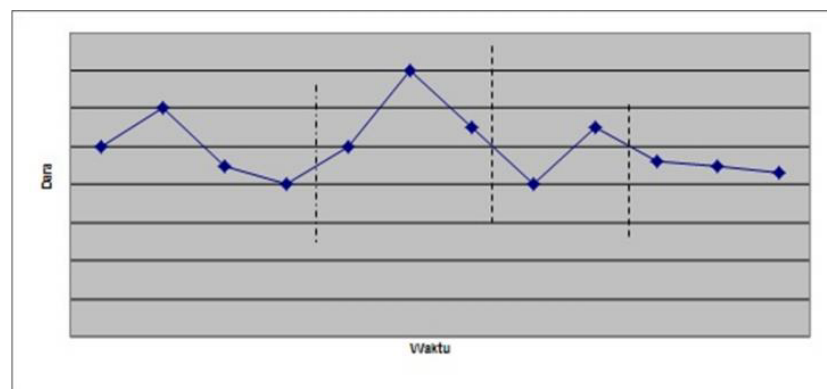
Pola data ini terjadi bila nilai data berfluktuasi di sekitar rata-rata yang konstan membentuk garis horisontal.



Gambar 1. Pola data horisontal

2. Pola Data Musiman (*Seasonal*)

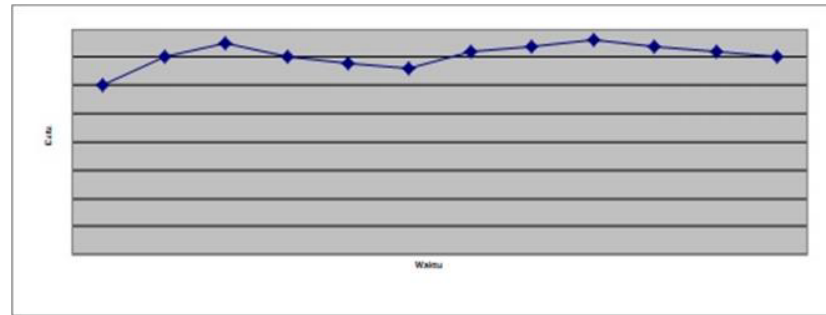
Pola yang menunjukkan perubahan yang berulang-ulang secara periodik dalam deret waktu. Pola yang terjadi bila suatu deret dipengaruhi oleh faktor musiman misalnya kuartal tahun tertentu, bulan, atau hari-hari pada minggu tertentu.



Gambar 2. Pola data musiman

3. Pola Data Siklus (*Cyclical*)

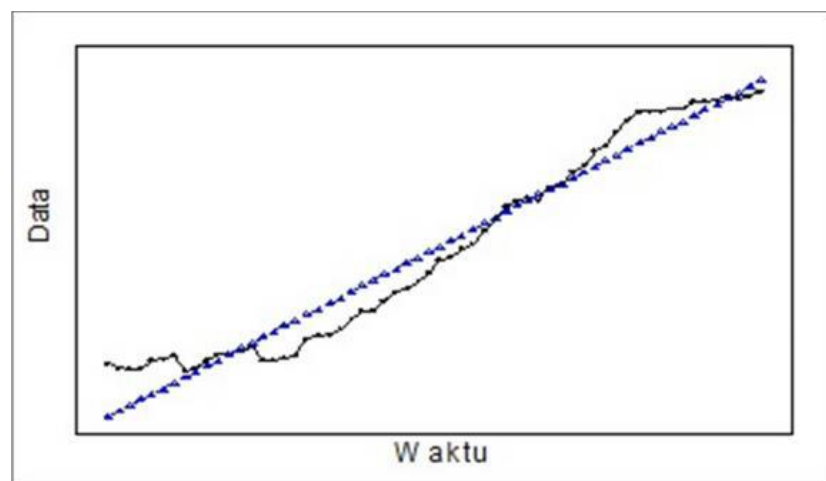
Pola data yang menunjukkan gerakan naik atau turun dalam jangka panjang dari suatu kurva *trend*, yang terjadi akibat pengaruh fluktuasi ekonomi jangka panjang seperti yang berhubungan dengan siklus bisnis.



Gambar 3. Pola data siklus

4. Pola Data *Trend*

Pola data *trend* terjadi saat terdapat kenaikan atau penurunan selama periode jangka panjang dalam data



Gambar 4. Pola data *trend*

2.5 Metode *k-Nearest Neighbor (k-NN) Regression*

Menurut Sinta (2014), *k-Nearest Neighbor (k-NN)* merupakan salah satu metode yang menggunakan algoritma tersupervisi dimana data pengujian diklasifikasikan berdasarkan kelas mayoritas. Tujuan dari algoritma ini adalah untuk mengklasifikasikan objek baru berdasarkan atribut dan data latih. Prinsip *k-NN* adalah menemukan k objek dari data latih yang paling dekat dengan data uji.

Algoritma *k-NN* sangat sederhana, yaitu bekerja berdasarkan jarak terdekat dari data uji ke data latih untuk menentukan k tetangga terdekatnya, kemudian diambil sebagian besar k tetangga terdekatnya untuk dijadikan prediksi dari data pengujian. Adapun cara untuk mengukur kedekatan antara data latih dengan data uji, diantaranya dengan menggunakan jarak *Euclidean* dan jarak *Manhattan*. Namun yang sering digunakan untuk mengukur kedekatan antardata adalah dengan menggunakan jarak *Euclidean*.

Metode *k-NN* merupakan salah satu algoritma *Machine Learning (ML)* yang juga suatu metode sederhana untuk diterapkan dalam analisis data dengan dimensi peubah yang banyak (Alkhatib, 2013). *k-NN* pada dasarnya digunakan untuk melakukan klasifikasi, tetapi dapat juga digunakan untuk melakukan regresi yang biasa disebut juga sebagai *k-NN Regression*.

Menurut Song dkk. (2017), regresi k -NN didasarkan pada pembelajaran dengan membandingkan contoh uji yang diberikan dengan set pelatihan. Misalkan $T = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$ menunjukkan set pelatihan dengan matriks jarak D , dimana $x_j = x_{j_1}, x_{j_2}, \dots, x_{j_m}$ adalah data ke- j yang dilambangkan dengan indeks m menghasilkan output y_j dan N adalah jumlah data. Ketika diberikan contoh uji pada x , perlu menghitung jarak D_j antara x dan setiap contoh x_j dalam T . Mengurutkan jarak D_j dengan nilainya, jika D_j diberi peringkat posisi ke- j maka jarak D_j contoh yang sesuai disebut sebagai j tetangga terdekat $NN_j(x)$ dan *output*-nya ditulis sebagai $y_j(x)$. Oleh karena itu, hasil prediksi dari x adalah rata-rata dari *output* k tetangga terdekatnya dalam regresi yaitu:

$$\hat{Y}_i = \frac{1}{k} \sum_{j=1}^k y_j(x) \quad (2.3)$$

dengan:

\hat{Y}_i = nilai prediksi,

k = banyaknya tetangga terdekat dari $y_j(x)$,

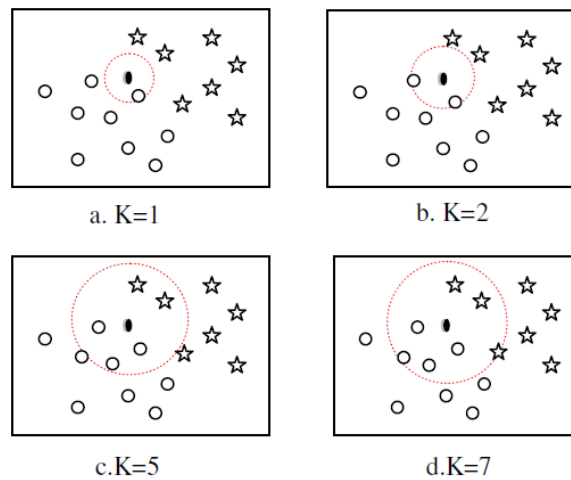
$y_j(x)$ = nilai data latih *output* sesuai urutan jarak berdasarkan jarak

Euclidean.

Algoritma *k-Nearest Neighbor* (*k*-NN) merupakan algoritma termasuk kedalam kategori *Distance-Based Algorithms* (Wang dkk., 2007). *Distance-Based Algorithms* adalah algoritma yang menentukan kemiripan data atau objek berdasarkan pada kedekatan jarak antar data ke suatu kelas atau kelompok data lainnya (Kataria dan Singh, 2013). Kemiripan antar data pada *k*-NN ditentukan dengan menggunakan pengukuran model jarak.

Metode *k*-NN bekerja dengan melihat *k* tetangga terdekat dari data, nilai *k* pada *k*-NN merupakan jumlah tetangga terdekat jika *k* bernilai 1, maka kelas dari satu data latih yang terdekat akan menjadi kelas bagi data uji yang baru. Jika *k* bernilai 3, maka akan diambil tiga data latih terdekat yang akan dijadikan kelas bagi data uji yang baru. Begitu juga jika nilai $k = 4, 5, 6, 7$, dan seterusnya (Lidya, 2014). Pemilihan nilai *k* yang kecil menyebabkan hasil klasifikasi sensitif terhadap *noise*, jika nilai *k* terlalu besar kemungkinan jumlah tetangga terdekat dari kelas lain terlalu banyak yang akhirnya dapat menurunkan hasil klasifikasi (Gou dkk., 2014).

Untuk lebih jelasnya perhatikan Gambar 5, ada dua kelas sebagai sampel yaitu lingkaran dan bintang, dimana oval yang berwarna hitam adalah data baru yang akan diklasifikasikan oleh algoritma k -NN (Syaliman, 2018).



Gambar 5. K -NN dengan nilai: (a) $K = 1$, (b) $K = 2$, (c) $K = 5$, (d) $K = 7$

Salah satu masalah yang dihadapi k -NN adalah dalam pemilihan nilai k yang tepat. Pemilihan nilai k yang besar dapat mengakibatkan distorsi data yang besar pula. Hal ini dikarenakan setiap tetangga mempunyai bobot yang sama terhadap data uji, sedangkan k yang terlalu kecil bisa menyebabkan algoritma terlalu sensitif terhadap *noise*.

Adapun langkah-langkah algoritma *k*-NN *Regression* untuk menghitung prediksi adalah sebagai berikut (Seruni dkk., 2020):

1. menentukan nilai *k* dan fungsi jarak. Nilai *k* merupakan jumlah kelompok terdekat yang ditentukan pengguna, yang digunakan untuk mengelompokkan data latih target yang dipilih berdasarkan jarak antara data latih dan data uji,
2. menghitung jarak antara data yang akan diprediksi dengan data-data sebelumnya,
3. mengurutkan data-data sebelumnya berdasarkan jarak terkecil, kemudian menghitung rata-rata target dari *k* data teratas yang dapat dilihat dari Persamaan 2.3. Hasil perhitungan rata-rata tersebut menjadi hasil prediksinya.

2.6 Mean Absolute Percentage Error (MAPE)

Menurut Makridakis dkk. (1999), MAPE digunakan untuk melakukan perhitungan perbedaan antara data aktual dan data hasil peramalan. Perbedaan tersebut diabsolutkan, kemudian dihitung ke dalam bentuk persentase terhadap data asli. Hasil persentase tersebut kemudian didapatkan nilai rata-ratanya. Suatu model mempunyai kinerja sangat bagus jika nilai MAPE berada di bawah 10% dan mempunyai kinerja bagus jika nilai MAPE berada diantara 10% dan 20%. MAPE dapat dihitung dengan menggunakan persamaan berikut:

$$MAPE = \frac{1}{n} \left(\left| \frac{y_1 - \hat{y}_1}{y_1} \right| + \left| \frac{y_2 - \hat{y}_2}{y_2} \right| + \left| \frac{y_3 - \hat{y}_3}{y_3} \right| + \dots + \left| \frac{y_n - \hat{y}_n}{y_n} \right| \right) \times 100\%$$

$$MAPE = \frac{1}{n} \left(\sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \right) \times 100\% \quad (2.4)$$

dengan:

$MAPE = \text{mean percentage error,}$

n = banyaknya data yang diamati,

\hat{Y}_i = peramalan ke- i ,

y_i = data latih *output* ke- i .

2.7 Indeks Pembangunan Manusia (IPM)

Menurut Badan Pusat Statistik (2008), pengukuran pembangunan manusia pertama kali diperkenalkan oleh United Nations Development Programme (UNDP) pada tahun 1990. UNDP memperkenalkan sebuah gagasan baru dalam pengukuran pembangunan manusia yang disebut sebagai Indeks Pembangunan Manusia (IPM). Sejak saat itu, IPM dipublikasikan secara berkala dalam laporan tahunan *Human Development Report* (HDR). IPM menjelaskan bagaimana penduduk dapat mengakses hasil pembangunan dalam memperoleh pendapatan, kesehatan, pendidikan, dan sebagainya.

Indeks Pembangunan Manusia mengukur capaian pembangunan manusia berbasis sejumlah komponen dasar kualitas hidup. Sebagai ukuran kualitas hidup, IPM dibangun melalui pendekatan tiga indeks dasar. Indeks tersebut mencakup:

1. umur panjang dan hidup sehat (*a long and healthy life*),
2. pengetahuan (*knowledge*), dan
3. standar hidup layak (*decent standard of living*).

Ketiga indeks dasar pembangun IPM memiliki pengertian yang sangat luas, hal tersebut dikarenakan masing-masing indeks memiliki keterkaitan dengan banyak faktor yang mempengaruhinya. Untuk mengukur indeks kesehatan, digunakan angka umur harapan hidup (AHH). Selanjutnya untuk mengukur indeks pendidikan digunakan gabungan indikator angka melek huruf (AMH) dan rata-rata lama sekolah (RLS). Adapun untuk mengukur indeks standar hidup layak digunakan indikator kemampuan daya beli yang dilihat melalui pengeluaran perkapita riil pertahun (PPP) yang disesuaikan (BPS, 2008).

2.7.1 Indeks Kesehatan

Menurut Badan Pusat Statistik (2008), angka harapan hidup (AHH) merupakan perkiraan banyak tahun yang dapat ditempuh oleh seseorang selama hidup. AHH dihitung menggunakan pendekatan tak langsung. Menghitung indeks kesehatan distandarisasi dengan nilai minimum dan maksimum sebelum digunakan untuk menghitung IPM, berikut rumus indeks kesehatan:

$$I_{kesehatan} = \frac{AHH - AHH_{min}}{AHH_{max} - AHH_{min}} \quad (2.5)$$

dengan:

- $I_{kesehatan}$ = indeks kesehatan,
- AHH = nilai angka harapan hidup,
- AHH_{min} = nilai angka harapan hidup terendah,
- AHH_{max} = nilai angka harapan hidup tertinggi.

2.7.2 Indeks Pendidikan

Indikator pendidikan yang mempresentasikan indeks pendidikan dalam IPM adalah angka melek huruf (AMH) dan rata-rata lama sekolah (RLS). Akan tetapi, indikator angka melek huruf diganti dengan indikator baru yang disebut harapan lama sekolah (HLS) karena angka melek huruf sudah tidak relevan lagi dengan kondisi saat ini sehingga diganti dengan harapan lama sekolah. Pada metode lama, cakupan penduduk yang dihitung adalah penduduk berusia 15 tahun ke atas. Sementara pada metode baru, cakupan penduduk yang dihitung berusia 25 tahun ke atas sesuai dengan rekomendasi UNDP (BPS, 2015).

Menghitung indeks pendidikan distandarisasi dengan nilai minimum dan maksimum sebelum digunakan untuk menghitung IPM, berikut rumus indeks pendidikan:

$$I_{HLS} = \frac{HLS - HLS_{min}}{HLS_{max} - HLS_{min}} \quad (2.6)$$

dengan:

- I_{HLS} = indeks harapan lama sekolah,
- HLS = nilai harapan lama sekolah,
- HLS_{min} = nilai harapan lama sekolah terendah,
- HLS_{max} = nilai harapan lama sekolah tertinggi.

$$I_{RLS} = \frac{RLS - RLS_{min}}{RLS_{max} - RLS_{min}} \quad (2.7)$$

dengan:

- I_{RLS} = indeks rata-rata lama sekolah,
 RLS = nilai rata-rata lama sekolah,
 RLS_{min} = nilai rata-rata lama sekolah terendah,
 RLS_{max} = nilai rata-rata lama sekolah tertinggi.

sehingga

$$I_{Pendidikan} = \frac{I_{HLS} + I_{RLS}}{2} \quad (2.8)$$

dengan:

- $I_{Pendidikan}$ = indeks pendidikan.

2.7.3 Indeks Standar Hidup Layak

Dimensi ketiga dari ukuran kualitas hidup manusia adalah standar hidup layak yang menggambarkan tingkat kesejahteraan penduduk sebagai dampak semakin membaiknya perekonomian. Badan Pusat Statistik menghitung standar rata-rata pengeluaran perkapita riil pertahun yang disesuaikan (PPP) dengan formula Atkinson (BPS, 2008). Menghitung indeks standar hidup layak distandarisasi dengan nilai minimum dan maksimum sebelum digunakan untuk menghitung IPM, berikut rumus indeks standar hidup layak:

$$I_{PPP} = \frac{PPP - PPP_{min}}{PPP_{max} - PPP_{min}} \quad (2.9)$$

dengan:

I_{PPP} = indeks pengeluaran perkapita riil pertahun,

PPP = nilai pengeluaran perkapita riil pertahun,

PPP_{min} = nilai pengeluaran perkapita riil pertahun terendah,

PPP_{max} = nilai pengeluaran perkapita riil pertahun tertinggi.

III. METODOLOGI PENELITIAN

3.1 Waktu dan Tempat Penelitian

Penelitian ini dilakukan pada semester genap tahun akademik 2019/2020, bertempat di Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.

3.2 Data Penelitian

Data yang akan digunakan dalam penelitian terdiri dari data Indeks Pembangunan Manusia (y), Angka Harapan Hidup (x_1), Rata-rata Lama Sekolah (x_2), Harapan Lama Sekolah (x_3), dan Pengeluaran Perkapita Riil Pertahun yang Disesuaikan (x_4). Data tersebut merupakan data sekunder yang diperoleh dengan mengakses *website* Badan Pusat Statistik (BPS) Indonesia dari tahun 2020-2021 sebanyak 34 provinsi di Indonesia, dimana data pada tahun 2020 merupakan data latih dan data pada tahun 2021 merupakan data uji.

3.3 Metode Penelitian

Untuk mempermudah hasil perhitungan dan hasil yang akurat penelitian ini menggunakan *software* RStudio 4.0.2. Adapun langkah-langkah penelitian menggunakan metode *k*-NN *Regression* yang akan dilakukan sebagai berikut.

1. Menentukan data latih dan data uji. Data IPM tahun 2020 dijadikan data latih dan data IPM tahun 2021 dijadikan data uji.
2. Menghitung jarak *Euclidean* ($D_{i,j}$) antara data latih dan data uji menggunakan Persamaan 2.2.
3. Mengurutkan jarak *Euclidean* ($D_{i,j}$) berdasarkan jarak terdekat hingga jarak terjauh.
4. Menentukan nilai k jumlah tetangga terdekat. Nilai k yang digunakan pada penelitian ini sesuai dengan jumlah data latih dan data uji yaitu 34.
5. Menghitung rata-rata (\hat{Y}_i) dari nilai objek pada jangkauan k dengan menggunakan kategori *k*-NN yang terdekat, sehingga dapat diprediksi nilai IPM untuk menghitung rata-rata pada Persamaan 2.3.
6. Melakukan evaluasi terhadap hasil akhir peramalan menggunakan *k*-NN *Regression* berdasarkan nilai MAPE dengan Persamaan 2.4, untuk menentukan hasil peramalan terbaik.
7. Analisis hasil peramalan dengan nilai MAPE terkecil.

V. KESIMPULAN

Berdasarkan hasil analisis dan pembahasan yang telah didapatkan pada bab sebelumnya dapat disimpulkan bahwa:

1. Metode *k*-NN *Regression* dapat digunakan untuk menyelesaikan prediksi indeks pembangunan manusia di Indonesia.
2. Berdasarkan analisis prediksi terbaik pada data IPM di Indonesia, nilai *k* yang optimal untuk data tersebut adalah $k = 2$ dengan galat MAPE 0,9019%.
3. Nilai prediksi IPM di Indonesia pada tahun 2022 menggunakan metode *k*-NN *Regression* diperoleh rata-rata sebesar 71,14.

DAFTAR PUSTAKA

- Alkhatib, K., Hassan, N., dan Ismail, H. 2013. Stock Price Prediction Using K-Nearest Neighbor (KNN) Algorithm. *International Journal of Business, Humanities and Technology*. **3**(3): 32-44.
- Arga, W. 1984. *Analisis Runtun Waktu Teori dan Aplikasi*. BPFE Yogyakarta, Yogyakarta.
- Badan Pusat Statistik. 2008. Indeks Pembangunan Manusia 2006-2007. *Katalog BPS: 4102002*. BPS, Jakarta.
- Badan Pusat Statistik. 2015. Indeks Pembangunan Manusia 2014 Metode Baru. *Katalog BPS: 4102002*. BPS, Jakarta.
- Fauzi, F. 2017. K-Nearest Neighbor (K-NN) dan Support Vector Machine (SVM) untuk Klasifikasi Indeks Pembangunan Manusia Provinsi Jawa Tengah. *Jurnal MIPA*. **40**(2): 118-124.
- Gou, J., Zhang, Y., Rao, Y., Shen, X., Wang, X., dan He, W. 2014. Improved Pseudo Nearest Neighbor Classification. *Knowledge-Based Systems*. **70**: 361-375.
- Kataria, A. dan Singh, M.D. 2013. A Review of Data Classification Using K-Nearest Neighbour Algorithm. *International Journal of Emerging Technology and Advanced Engineering*. **3**(6): 354-360.
- Lidya, S.K. 2014. Sentiment Analysis pada Teks Bahasa Indonesia Menggunakan Support Vector Machine (SVM) dan K-Nearest Neighbor (K-NN). Tesis. Jurusan Magister Teknik Informatika Fasilkom-TI USU, Medan.

- Makridakis, S., Steven, C.W., dan Victor, E.M. 1999. *Metode dan Aplikasi Peramalan*. Terjemahan Hari Suminto. Binarupa Aksara, Jakarta.
- Putra, D.M. 2015. *Pemodelan Indeks Pembangunan Manusia (IPM) Provinsi Jawa Timur dengan Menggunakan Metode Regresi Logistik Ridge*. Tugas Akhir. Jurusan Statistika FMIPA ITS, Surabaya.
- Ripley, B.D. 1996. *Pattern Recognition and Neural Networks*. Cambridge University Press, Cambridge.
- Rokach, L. 2010. *Patterns Classification Using Ensemble Methods*. World Scientific, Singapore.
- Santosa, B. 2007. *Data Mining: Teknik Pemanfaatan Data Untuk Keperluan Bisnis*. Graha Ilmu, Yogyakarta.
- Seruni, D.S., Muhammad, T.F., dan Randy, C.W. 2020. Sistem Prediksi Pertumbuhan Jumlah Penduduk Kota Malang Menggunakan Metode K-Nearest Neighbor Regression. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*. 4(4): 1075-1082.
- Sinta, D. 2014. Ensemble K-Nearest Neighbors Method to Predict Rice in Indonesia. *Applied Mathematical Sciences*. 8(160): 7993-8005.
- Song, Y., Liang, J., Lu, J., dan Zhao, X. 2017. An Efficient Instance Selection Algorithm for k Nearest Neighbor Regression. *Neurocomputing*. 251: 26-34.
- Subagyo, P. 1986. *Forecasting Konsep dan Aplikasi*. BPFE Yogyakarta, Yogyakarta.
- Syaliman, K.U. 2018. Peningkatan Akurasi pada Metode Klasifikasi K-Nearest Neighbor Menggunakan Local Mean Based dan Distance Weight K-Nearest Neighbor. Tesis. Jurusan Magister Teknik Informatika Fasilkom-TI USU, Medan.

Taufiqurrohman, Achmad, H., dan Ajob, A. 2013. Pengenalan Plat Nomor Sepeda Motor dengan Menggunakan Metode Jarak Euclidean. *Transient*. **2**(3): 573-580.

Wang, J., Neskovic, P., dan Cooper, L.N. 2007. Improving Nearest Neighbor Rule with A Simple Adaptive Distance Measure. *Pattern Recognition Letter*. **28**: 207-213.