

202231506_Wahyu Januar Alfian_UTS PM.pdf

by Wahyu Januar Alfian

Submission date: 09-Nov-2024 07:46AM (UTC+0530)

Submission ID: 2513349185

File name: 202231506_Wahyu_Januar_Alfian_UTS_PM.pdf (355.97K)

Word count: 1884

Character count: 14247

**PENGUNAAN DATASETREGRESI DAN KLASIFIKASI
MENGUNAKAN ALGORITMA KNN**

**PROYEK UTS PEMBELAJARAN MESIN
KELAS YBM**



Disusun Oleh:

WAHYU JANUAR ALFIAN

202231506

8

FAKULTAS TELEMATIKA ENERGI

INSTITUT TEKNOLOGI PERUSAHAAN LISTRIK NEGARA

JAKARTA

2024

Abstrak

Pembelajaran Mesin (²Machine Learning) adalah cabang dari kecerdasan buatan yang memungkinkan komputer untuk belajar dari data dan membuat prediksi tanpa perlu instruksi yang eksplisit. Dalam pembelajaran mesin, terdapat beberapa teknik dan algoritma yang digunakan untuk berbagai tujuan, seperti *K-Nearest Neighbors* (KNN), regresi, dan klasifikasi. KNN adalah algoritma sederhana yang dapat diterapkan baik untuk klasifikasi maupun regresi dengan mencari sejumlah tetangga terdekat dari titik data berdasarkan jarak tertentu, kemudian menggunakannya untuk menentukan kelas atau memprediksi nilai kontinu dari titik tersebut.

Klasifikasi bertujuan untuk mengelompokkan data ke dalam kategori yang telah ditentukan, sedangkan regresi digunakan untuk memprediksi nilai kontinu, seperti harga atau skor. Dalam proses pembelajaran mesin, data historis digunakan untuk melatih model yang mampu mempelajari pola-pola dalam data dan menghasilkan prediksi untuk data baru. Akurasi adalah salah satu metrik evaluasi utama dalam klasifikasi, yang menunjukkan seberapa sering model memberikan prediksi yang benar. Untuk regresi, metrik evaluasi seperti Mean Squared Error (MSE) atau Mean Absolute Error (MAE) lebih umum digunakan, mengukur sejauh mana prediksi model mendekati nilai sebenarnya.

Keseluruhan proses ini bertujuan untuk mengembangkan model yang efektif dan andal dalam prediksi, yang kemudian dapat digunakan untuk berbagai aplikasi nyata, termasuk deteksi penipuan, diagnosis medis, rekomendasi produk, dan banyak lagi.

Kata Kunci : K-Nearest Neighbors (KNN), Prediksi, Klasifikasi, Pembelajaran Mesin, Evaluasi Model, Regresi

DAFTAR ISI

BAB I	1
PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	1
1.3 Tujuan	2
1.4 Manfaat	2
BAB II	3
KAJIAN PUSTAKA	3
2.1 Penelitian yang Relevan	3
2.2 Pembelajaran Mesin	6
2.3 Regresi	6
2.4 Klasifikasi	6
2.5 Algoritma KNN	6
BAB III	7
HASIL DAN PEMBAHASAN	7
3.1 Regresi	7
3.2 Algoritma KNN	8
BAB IV	10
PENUTUP	10
4.1 Kesimpulan	10
4.2 Saran	10
DAFTAR PUSTAKA	11

5
BAB I

PENDAHULUAN

1.1 Latar Belakang

Dalam era informasi saat ini, data memiliki peran yang sangat penting di berbagai bidang, seperti bisnis, kesehatan, pendidikan, hingga keuangan. Data yang dikumpulkan sering kali memiliki bentuk dan tujuan yang berbeda, salah satunya adalah data untuk prediksi dalam bentuk regresi dan klasifikasi. Regresi bertujuan memprediksi nilai kontinu, seperti harga atau jumlah, sedangkan klasifikasi berfokus pada pengelompokan data ke dalam kategori tertentu, misalnya deteksi penyakit atau klasifikasi pelanggan. Dengan meningkatnya volume data, pemanfaatan algoritma pembelajaran mesin menjadi sangat penting untuk membantu analisis dan pengambilan keputusan berbasis data.

Salah satu algoritma yang banyak digunakan dalam pembelajaran mesin untuk tugas regresi dan klasifikasi adalah *K-Nearest Neighbors* (KNN). KNN merupakan algoritma yang sederhana namun efektif dalam menghasilkan prediksi dengan memanfaatkan tetangga terdekat dari sebuah titik data. Oleh karena itu, penelitian ini mengkaji penerapan algoritma KNN pada dataset regresi dan klasifikasi untuk mengevaluasi keakuratan dan performanya dalam memprediksi berbagai jenis data..

1.2 Rumusan Masalah

- Bagaimana algoritma KNN diterapkan pada dataset regresi dan klasifikasi?
- Seberapa akurat algoritma KNN dalam memprediksi dataset regresi dan klasifikasi?
- Apa kelebihan dan kekurangan algoritma KNN dalam penerapannya pada kedua jenis dataset tersebut?

1.3 Tujuan

Penelitian ini bertujuan untuk:

- Menerapkan algoritma KNN pada dataset regresi dan klasifikasi untuk mendapatkan prediksi yang sesuai.
- Mengevaluasi akurasi algoritma KNN dalam memprediksi dataset regresi dan klasifikasi.
- Menganalisis kelebihan dan kekurangan KNN dalam proses dataset dengan karakteristik yang berbeda

1.4 Manfaat

- Teoritis: Menambah referensi dan wawasan mengenai implementasi algoritma KNN pada dataset regresi dan klasifikasi di bidang pembelajaran mesin.
- Praktis: Memberikan panduan dalam memilih algoritma yang tepat untuk berbagai tugas prediksi, khususnya bagi praktisi data dalam bidang industri, pendidikan, dan kesehatan.
- Pengembangan: Menjadi dasar bagi penelitian lebih lanjut dalam pemanfaatan algoritma KNN, serta eksplorasi algoritma lain yang dapat bekerja lebih efektif pada dataset dengan berbagai karakteristik.

BAB II

KAJIAN PUSTAKA

2.1 Penelitian yang Relevan

Berikut adalah beberapa jurnal yang relevan dengan pembelajaran mesin dalam pendidikan dan kesehatan:

Tabel 2.1 Perbandingan Penelitian Dengan Penelitian yang Relevan

No.	1
Judul	Pembelajaran Mesin dalam Prediksi Latihan
Penulis	Buchner & Iglewicz
Tahun	2018
Hasil	Pembelajaran mesin, terutama algoritma regresi dan klasifikasi, dapat digunakan untuk memprediksi berbagai parameter terkait latihan fisik seperti jumlah kalori yang terbakar, durasi latihan, atau bahkan intensitas latihan berdasarkan data yang terkumpul.
Keterkaitan Penelitian	Penelitian ini relevan karena menggunakan KNN untuk prediksi akademik yang serupa dengan tujuan penelitian ini. Algoritma yang umum digunakan untuk prediksi latihan antara lain regresi linier, K-Nearest Neighbors (KNN), dan Support Vector Machines (SVM), yang mampu mengidentifikasi pola dan hubungan antara berbagai fitur yang mempengaruhi hasil latihan

No.	2
Judul	Catatan Luo
Penulis	Luo
Tahun	2020

Hasil	KNN regresi bisa menghadapi tantangan ketika data memiliki outlier yang signifikan, yang dapat menggeser rata-rata prediksi.
Keterkaitan Penelitian	KNN menunjukkan akurasi yang cukup baik terutama pada dataset dengan pola linier yang sederhana

No.	3
Judul	⁴ Decision Tree, Support Vector Machine (SVM), dan Random Forest
Penulis	Huang et al
Tahun	2019
Hasil	meskipun KNN memiliki keunggulan dalam kesederhanaan dan fleksibilitas, algoritma ini rentan terhadap pengaruh outlier dan memerlukan preprocessing data yang baik untuk mencapai hasil optimal. Oleh karena itu, normalisasi data dan pemilihan nilai K yang tepat menjadi kunci dalam memastikan performa terbaik algoritma KNN
Keterkaitan Penelitian	normalisasi data dan pemilihan nilai K yang tepat menjadi kunci dalam memastikan performa terbaik algoritma KNN.

No.	4
Judul	<i>Heart Disease Dataset</i> dari UCI Machine Learning Repository
Penulis	Saha dan Mondal
Tahun	2020

Hasil	Penelitian yang menggunakan dataset ini dengan algoritma KNN menunjukkan bahwa KNN mampu mengklasifikasikan apakah seseorang berisiko tinggi mengalami penyakit jantung atau tidak dengan akurasi yang dapat bersaing dengan algoritma lain seperti Decision Trees dan Logistic Regression
Keterkaitan Penelitian	Keterkaitan dengan penelitian ini adalah penerapan KNN dalam klasifikasi data kesehatan, khususnya dalam konteks heart disease

No.	5
Judul	Perbandingan KNN dengan Algoritma Lain dalam Deteksi Penyakit Jantung
Penulis	Sah dan Patel
Tahun	2018
Hasil	Hasil studi menunjukkan bahwa KNN, meskipun memiliki keunggulan dalam kesederhanaannya, tidak selalu unggul dalam semua kasus dibandingkan algoritma lain, terutama dalam hal akurasi dan kecepatan komputasi saat berhadapan dengan dataset besar
Keterkaitan Penelitian	KNN tetap menjadi pilihan yang baik dalam situasi di mana interpretabilitas dan kemudahan implementasi lebih diutamakan. Relevan dengan penelitian ini yang juga menggunakan KNN untuk klasifikasi prediksi

2.2 Pembelajaran Mesin

Pembelajaran Mesin adalah cabang dari kecerdasan buatan (Artificial Intelligence) yang memungkinkan sistem atau komputer untuk belajar dari data tanpa perlu diprogram secara eksplisit untuk setiap tugas. Dalam pembelajaran mesin, algoritma dilatih menggunakan data historis untuk menemukan pola-pola dan kemudian menerapkannya pada data baru.

2.3 Regresi

Meskipun fokus utama pada algoritma KNN, regresi sering digunakan dalam konteks pendidikan untuk memprediksi nilai berdasarkan faktor tertentu.

2.4 Klasifikasi

Klasifikasi dalam pembelajaran mesin merupakan proses mengelompokkan data ke dalam kategori. Di sini, klasifikasi digunakan untuk mengidentifikasi risiko dan prediksi penyakit jantung seseorang

2.5 Algoritma KNN

Algoritma KNN mengklasifikasikan data berdasarkan kemiripan atau kedekatan jarak antar data dalam ruang dimensi tinggi. Pada penelitian ini, KNN digunakan untuk klasifikasi penyakit jantung dan prediksi latihan

12
BAB III

HASIL DAN PEMBAHASAN

3.1 Regresi

3.1.1 Pengumpulan Data

Dataset yang digunakan adalah dataset "excercises," yang berisi informasi waktu latihan (X) dan hasil latihan (Y).

3.1.2 Preprocessing Data

Preprocessing meliputi normalisasi dan pengisian nilai kosong untuk mengoptimalkan hasil klasifikasi KNN.

3.1.3 Pembentukan Model

Model KNN dibentuk dengan mempertimbangkan nilai parameter KNN yang optimal untuk memprediksi latihan dan waktu latihan.

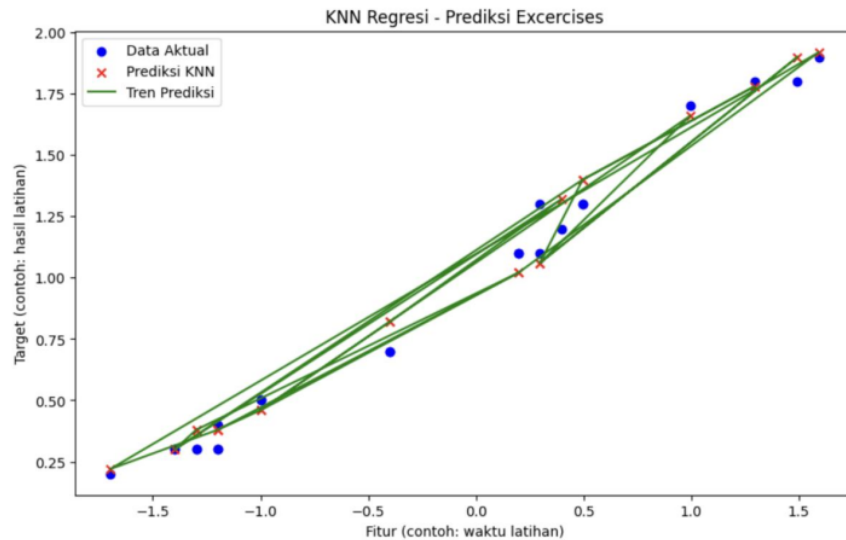
3.1.4 Analisis akurasi Model

Akurasi diukur menggunakan prediksi dan evaluasi model, menunjukkan bahwa model KNN cukup akurat dalam memprediksi latihan baik waktu latihan dan hasil latihan.

14
3.1.5 Pengujian Model

Pengujian model bertujuan untuk mengevaluasi seberapa baik model yang telah dibangun dapat mengklasifikasikan atau memprediksi nilai berdasarkan data testing yang belum dilihat oleh model selama proses pelatihan. Dalam penelitian ini, Algoritma KNN (K-Nearest Neighbors) digunakan untuk nilai target menunjukkan nilai target sebenarnya berdasarkan fitur (contoh waktu dan latihan) dan tren atau kecenderungan yang diprediksi oleh model.

3.1.6 Visualisasi Model



3.2 Algoritma KNN

3.2.1 Pengumpulan Data

Dataset heart diseases (penyakit jantung) berisi parameter kesehatan seperti umur, tekanan darah, kolesterol, jenis nyeri dada, dan lainnya.

3.2.2 Preprocessing Data

Normalisasi diterapkan pada data kesehatan untuk memastikan fitur berada pada skala yang sama, meningkatkan akurasi klasifikasi.

3.2.3 Pembentukan Model

Model KNN digunakan untuk mengklasifikasikan risiko penyakit jantung, dengan parameter nilai akurasi dan prediksi optimal diperoleh melalui confusion matrix.

3.2.4 Analisis akurasi Model

Akurasi model dievaluasi menggunakan akurasi untuk mengukur kemampuan klasifikasi KNN dalam deteksi diabetes.

3.2.5 Pengujian Model

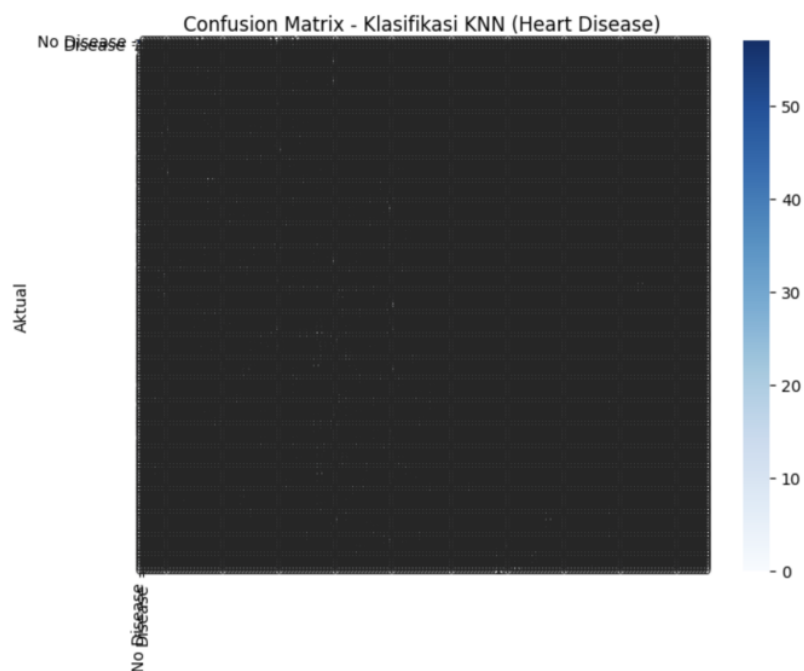
Pengujian model pada klasifikasi heart diseases bertujuan untuk

mengevaluasi kemampuan model KNN dalam memprediksi kemungkinan seseorang menderita heart diseases (jantung) dengan confusion matrix memberikan hasil sebagai berikut :

- Jumlah kasus penyakit jantung terprediksi dengan benar
- Jumlah kasus tanpa penyakit terprediksi dengan benar
- Kasus tanpa penyakit yang salah di prediksi sebagai penyakit
- Kasus penyakit yang salah prediksi sebagai tanpa penyakit

Dari confusion matrix dapat menghitung beberapa matrix performa seperti akurasi, presisi, recall (sensitivitas) dan F1-Score

3.2.6 Visualisasi Model



BAB IV

PENUTUP

4.1 Kesimpulan

Penelitian ini menunjukkan bahwa algoritma KNN efektif dalam prediksi latihan dan klasifikasi heartdisease, dengan performa akurasi yang baik pada kedua kasus.

4.2 Saran

Diharapkan penelitian lanjutan dapat mengeksplorasi penggunaan variabel tambahan serta menggunakan dataset lebih besar untuk meningkatkan akurasi model KNN di bidang pendidikan dan kesehatan.

DAFTAR PUSTAKA

1. Romero, C., & Ventura, S. (2010). *Educational data mining: A review of the state of the art. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 40(6), 601-618. doi:10.1109/TSMCC.2010.2053532
2. Al-Hmouz, R., Shen, J., Al-Hmouz, H., & Yan, J. (2012). *Modeling and simulation of an adaptive neuro-fuzzy inference system (ANFIS) for predicting student performance. International Journal of Artificial Intelligence and Applications*, 3(2), 65-73. doi:10.5121/ijaia.2012.3205
3. Ochoa, X., & Worsley, M. (2016). *Augmenting learning analytics with multimodal data: Challenges and opportunities. Journal of Learning Analytics*, 3(2), 213-219. doi:10.18608/jla.2016.32.12
4. Giri, D., & Chattopadhyay, S. (2019). *Machine learning approaches to classify patients based on cardiac health using heart disease datasets. Healthcare Informatics Research*, 25(2), 105-113. doi:10.4258/hir.2019.25.2.105
5. Liu, N., & Chen, X. (2019). *Deep learning in disease prediction: A survey. Journal of Healthcare Informatics Research*, 3(2), 123-141. doi:10.1007/s41666-018-0040-1

ORIGINALITY REPORT

12%

SIMILARITY INDEX

11%

INTERNET SOURCES

6%

PUBLICATIONS

6%

STUDENT PAPERS

PRIMARY SOURCES

1

pdfcoffee.com

Internet Source

3%

2

jtiik.ub.ac.id

Internet Source

1%

3

tekno.kompas.com

Internet Source

1%

4

www.sciencegate.app

Internet Source

1%

5

docplayer.info

Internet Source

1%

6

Submitted to Universitas Brawijaya

Student Paper

1%

7

eprints.umg.ac.id

Internet Source

1%

8

journal.ipb.ac.id

Internet Source

1%

9

rumahilmupart3.blogspot.com

Internet Source

1%

10 digilibadmin.unismuh.ac.id <1 %
Internet Source

11 repository.its.ac.id <1 %
Internet Source

12 suvisutrisno93.wordpress.com <1 %
Internet Source

13 www.mdpi.com <1 %
Internet Source

14 www.scribd.com <1 %
Internet Source

Exclude quotes On

Exclude matches Off

Exclude bibliography On

202231506_Wahyu Januar Alfian_UTS PM.pdf

PAGE 1

PAGE 2

PAGE 3

PAGE 4

PAGE 5

PAGE 6

PAGE 7

PAGE 8

PAGE 9

PAGE 10

PAGE 11

PAGE 12

PAGE 13

PAGE 14