

ANALISIS KOMPARASI MODEL ALGORITMA SUPPORT VECTOR MACHINE DAN K-NEAREST NEIGHBORS PADA KLASIFIKASI KUALITAS UDARA KOTA JAKARTA

Oleh:

Nama: Yudha Aryadi Sani

NIM: 41200210

Skripsi

**Diajukan sebagai salah satu syarat
untuk memperoleh gelar Sarjana Komputer**

Program Studi Sistem Informasi



KWIK KIAN GIE
SCHOOL OF BUSINESS

INSTITUT BISNIS dan INFORMATIKA KWIK KIAN GIE

JAKARTA

Maret 2024

(C) Hak cipta milik IBI KKG (Institut Bisnis dan Informatika Kwik Kian Gie)

Hak Cipta Dilindungi Undang-Undang

Institut Bisnis dan Informatika Kwik Kian Gie

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik dan tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IBIKKG.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IBIKKG.



(C) Hak cipta milik IBI KKG (Institut Bisnis dan Informatika Kwik Kian Gie)

Institut Bisnis dan Informatika Kwik Kian Gie

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik dan tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IBIKKG.
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IBIKKG.

PENGESAHAN

JUDUL

ANALISIS KOMPARASI MODEL ALGORITMA SUPPORT VECTOR MACHINE DAN K-NEAREST NEIGHBORS PADA KLASIFIKASI KUALITAS UDARA KOTA JAKARTA

Diajukan Oleh:

Nama: Yudha Aryadi Sani

NIM: 41200210

Jakarta, 27 Maret 2024

Disetujui oleh:

Pembimbing

(Budi Wasito, S.Kom., M.M., M.Kom)

INSTITUT BISNIS DAN INFORMATIKA KWIK KIAN GIE

JAKARTA 2024

KATA PENGANTAR

Pertama-tama, marilah kita panjatkan puji syukur kami kepada Tuhan Yang Maha Esa atas rahmat dan karunia-Nya sehingga peneliti dapat menyelesaikan proposal skripsi dengan judul “Analisis Komparasi model Algoritma Support Vector Machine dan K-Nearest Neighbour Pada Klasifikasi Kualitas Udara Kota Jakarta” dengan tepat waktu.

Dengan ilmu dan pembelajaran yang telah disampaikan oleh para dosen, peneliti telah memperoleh banyak pengetahuan untuk menyusun karya akhir yang mencakup analisis komparasi algoritma data mining secara maksimal, yang merupakan syarat untuk memperoleh gelar Sarjana Komputer.

Peneliti telah menerima bantuan dari banyak orang yang telah membantu mereka menyelesaikan skripsi ini. Oleh sebab itu, penulis ingin mengucapkan terima kasih yang sebesar-besarnya kepada:

1. Bapak Joko Susilo, S.Kom, M.M., M.Kom. selaku kepala program studi Sistem Informasi dan Teknik Informatika yang memberikan arahan dari awal hingga akhir penulisan karya akhir.
2. Bapak Budi Wasito, S.Kom., M.M., M.Kom. selaku dosen pembimbing yang memberikan arahan dari awal hingga akhir penulisan karya akhir.
3. Para dosen Institut Bisnis dan Informatika Kwik Kian Gie khususnya dalam program studi Sistem Informasi dan Teknik Informatika yang telah membina dan membekali ilmu selama perkuliahan.
4. Keluarga yang selalu memberikan dukungan dan motivasi selama penulisan skripsi.
5. Teman-teman pada kampus Institut Bisnis dan Informatika Kwik Kian Gie terutama pada jurusan Sistem Informasi dan Teknik Informatika angkatan 2020 yang memberikan semangat dan dukungan secara fisik maupun mental.



Peneliti menyadari bahwa penelitian ini masih jauh dari kata sempurna dan memiliki banyak kelemahan. Jadi, peneliti sangat mengharapkan kritik dan saran yang membangun dari semua orang untuk membantu memperbaiki skripsi ini.

Akhir kata, peneliti mengucapkan terimakasih kepada pihak yang tidak dapat disebutkan yang telah memberikan banyak dukungan dan semangat kepada peneliti.

Jakarta, 27 Maret 2024

Yudha Aryadi Sani

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik dan tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IBIKKG.
2. Dilarang menggunakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IBIKKG.



ABSTRAK

Yudha Aryadi Sani / 41200210 / Analisis Komparasi Model Support Vector Machine dan K-Nearest Neighbor Pada Klasifikasi Kualitas Udara Kota Jakarta / Pembimbing: Budi Wasito, S.Kom, M.M., M.Kom.

Polusi udara telah menjadi salah satu tantangan lingkungan terbesar di seluruh dunia, dan tidak terkecuali bagi Kota Jakarta sebagai ibu kota Indonesia. Jakarta, sebagai salah satu kota terpadat di dunia, menghadapi masalah serius dalam memastikan kualitas udara yang sehat bagi penduduknya. Dampak dari polusi udara ini tidak hanya terbatas pada kesehatan manusia, tetapi juga merusak lingkungan secara keseluruhan, termasuk tanaman dan ekosistem air. Dalam upaya untuk mengatasi masalah ini, banyak penelitian telah dilakukan untuk mengembangkan model prediksi yang dapat memberikan perkiraan tingkat kualitas udara di Kota Jakarta. Dua model pemodelan yang dapat digunakan dalam konteks ini adalah Support Vector Machine (SVM) dan K-Nearest Neighbors (KNN). Penelitian ini bertujuan untuk membandingkan kinerja model algoritma SVM dengan algoritma KNN dalam klasifikasi kualitas udara Kota Jakarta.

Data mining merupakan seni dan ilmu untuk menemukan pengetahuan, wawasan, dan pola dalam data berdasarkan metodologi CRISP -DM (*Cross Industry Standard Process For Data Mining*)

Sumber data dalam penelitian ini adalah data indeks standar pencemaran udara (ISPU) Kota Jakarta yang didapatkan pada *website* satudata.jakarta.go.id

Penelitian ini membandingkan akurasi klasifikasi kualitas udara Kota Jakarta pada algoritma Support vector Machine(SVM) dan K-Nearest Neighbor (KNN) menggunakan python. Hasil Komparasi ditentukan dengan tingkat akurasi dan *score* dari *confusion matrix*. Hasil klasifikasi disajikan dalam bentuk *Graphic User Interface* dengan media antarmuka.

Hasil perbandingan algoritma klasifikasi pada dua model dalam data mining menunjukkan bahwa algoritma Support Vector Machine (SVM) lebih unggul dari K-Nearest Neighbor (KNN). Hal ini terbukti dari tingkat akurasi yang lebih tinggi pada SVM, khususnya dengan penggunaan kernel Rbf yang mencapai 97,05%, dibandingkan dengan KNN yang memiliki akurasi sebesar 94,74% dengan parameter $p=1$ dan $k=5$. Selain itu, SVM juga menunjukkan nilai prediksi benar yang lebih tinggi dibandingkan dengan KNN. Dalam rentang waktu 1 tahun secara keseluruhan mungkin cukup baik, dengan kualitas sedang di angka lebih dari sama dengan 50 sampai 100. Namun, tetap diperlukan pengendalian untuk daerah kategori tidak sehat.

Kata Kunci: Klasifikasi, *Data Mining*, *Support Vector Machine*, *K-Nearest Neighbor*, *Python*, Akurasi, *Confusion Matrix*, Pencemaran Udara, ISPU, CRISP -DM.



ABSTRACT

Yudha Aryadi Sani / 41200210 / Comparative Analysis of Support Vector Machine and K-Nearest Neighbor Models on Jakarta City Air Quality Classification / Advisor: Budi Wasito, S.Kom, M.M., M.Kom.

Air pollution has become one of the biggest environmental challenges worldwide, and Jakarta, the capital city of Indonesia, is no exception. Jakarta, as one of the most populous cities in the world, faces serious problems in ensuring healthy air quality for its residents. The impact of this air pollution is not only limited to human health, but also damages the environment as a whole, including plants and aquatic ecosystems. In an effort to address this issue, many studies have been conducted to develop prediction models that can provide estimates of air quality levels in Jakarta City. Two modeling models that can be used in this context are Support Vector Machine (SVM) and K-Nearest Neighbors (KNN). This study aims to compare the performance of the SVM algorithm model with the KNN algorithm in the classification of air quality in Jakarta City.

Data mining is the art and science of discovering knowledge, insights, and patterns in data based on the CRISP -DM (Cross Industry Standard Process For Data Mining) methodology.

The data source in this study is the Jakarta air pollution standard index (ISPU) data obtained on the satudata.jakarta.go.id website.

This study compares the accuracy of the Jakarta City air quality classification on the Support vector Machine (SVM) and K-Nearest Neighbor (KNN) algorithms using python. Comparison results are determined by the level of accuracy and score of the confusion matrix. Classification results are presented in the form of a Graphic User Interface with media interface.

The comparison of classification algorithms on two models in data mining shows that the Support Vector Machine (SVM) algorithm is superior to K-Nearest Neighbor (KNN). This is evident from the higher accuracy rate in SVM, especially with the use of the Rbf kernel which reaches 97.05%, compared to KNN which has an accuracy of 94.74% with parameters $p = 1$ and $k = 5$. In addition, SVM also shows a higher correct prediction value compared to KNN. In a 1-year time span the overall quality may be quite good, with moderate quality at more than equal to 50 to 100. However, control is still needed for the unhealthy category areas.

Keywords: Classification, Data Mining, Support Vector Machine, K-Nearest Neighbor, Python, Accuracy, Confusion Matrix, Air Pollution, ISPU, CRISP-DM.



DAFTAR ISI

PENGESAHAN.....	ii
KATA PENGANTAR.....	iii
ABSTRAK.....	v
ABSTRACT.....	vi
DAFTAR ISI.....	vii
DAFTAR TABEL.....	ix
DAFTAR GAMBAR.....	x
DAFTAR LAMPIRAN.....	xi
BAB I PENDAHULUAN.....	1
A. Latar Belakang Masalah.....	1
B. Identifikasi Masalah.....	2
C. Batasan Masalah.....	3
D. Tujuan Penelitian.....	3
E. Manfaat Penelitian.....	3
BAB II LANDASAN TEORI.....	4
A. Data.....	4
B. Informasi.....	7
C. Data Mining.....	8
D. Machine Learning.....	12
E. Klasifikasi.....	13
F. Sirkulus Penambangan Data CRISP-DM.....	13
G. Pencemaran udara.....	16
H. Algoritma Support Vector Machine (SVM).....	18
I. Algoritma K-Nearest Neighbors.....	19
J. Python.....	21
K. Penelitian Kuantitatif.....	21
L. Penelitian Terdahulu.....	21
BAB III METODOLOGI PENELITIAN.....	25
A. Gambaran Umum Objek Penelitian.....	25
B. Metodologi Penelitian.....	25
1. Teknik Pengumpulan Data.....	25
2. Teknik Analisis Data.....	26
3. Penerapan Algoritma.....	27
4. Teknik Pengukuran Data.....	29
C. Teknik Perancangan Graphic User Interface (GUI).....	34
BAB IV HASIL PENELITIAN DAN PEMBAHASAN.....	36
A. Architectural algorithm.....	36
1. Support Vector Machine Architecture.....	36



2. K-Nearest Neighbor Architecture.....	37
B. Exploratory Data Analysis (EDA).....	38
1. Praproses data.....	38
2. Pemisahan Data.....	41
C. Model Creation.....	43
1. Support Vector Machine (SVM).....	43
2. K-Nearest Neighbor (KNN).....	51
D. Komparasi Hasil Evaluasi dan Scoring.....	57
E. Rancangan Graphic User Interface (GUI).....	58
1. Home.....	58
2. Comparison Accuracy.....	59
3. Comparison Confusion Matrix.....	60
4. Resource.....	60
BAB V SIMPULAN DAN SARAN.....	62
A. Kesimpulan.....	62
B. Saran.....	63
DAFTAR PUSTAKA.....	64
LAMPIRAN.....	66
Lampiran 1 - Raw Dataset.....	66
Lampiran 2 - Praproses Data.....	75
Lampiran 3 - Pemisahan Data.....	76
Lampiran 4 - Model Creation.....	77

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik dan tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar IBIKKG.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IBIKKG.

DAFTAR TABEL

Tabel 4.1 Kategori Angka Rentang ISPU.....	33
Tabel 4.2 Penjelasan Nilai ISPU.....	34
Tabel 4.3 Hasil Evaluasi Matriks Model SVM.....	47
Tabel 4.4 Hasil Evaluasi Matriks Model KNN.....	53
Tabel 4.5 Hasil Komparasi Scoring.....	57

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik dan tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IBIKKG.
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IBIKKG.



DAFTAR GAMBAR



Hak cipta ini dilindungi undang-undang. Ulangan atau penjiplakan tanpa izin IBIKKG.

Institut Bisnis dan Informatika Kwik Kian Gie

Gambar 2.1	Proses data mining.....	9
Gambar 2.2	CRISP - DM data mining cycle.....	14
Gambar 3.1	Workflow Penerapan Algoritma.....	28
Gambar 3.2	Confusion Matrix.....	30
Gambar 3.3	Hierarki Rancangan GUI.....	35
Gambar 4.1	Komparasi Rata-rata Akurasi Setiap Kernel Model SVM.....	48
Gambar 4.2	Confusion Matrix Model SVM.....	50
Gambar 4.3	Komparasi Akurasi Nilai K dan P.....	54
Gambar 4.4	Confusion Matrix Model KNN.....	56
Gambar 4.5	Hasil Komparasi Confusion Matrix.....	57
Gambar 4.6	Tampilan Menu Home.....	58
Gambar 4.7	Tampilan Menu Comparison Accuracy.....	59
Gambar 4.8	Tampilan Menu Comparison Confusion Matrix.....	60
Gambar 4.9	Tampilan Menu Resource.....	61

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik dan tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IBIKKG.
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IBIKKG.

DAFTAR LAMPIRAN

Lampiran 1 - Raw Dataset.....	66
Lampiran 2 - Praproses Data.....	75
Lampiran 3 - Pemisahan Data.....	76
Lampiran 4 - Model Creation.....	77

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik dan tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IBIKKG.
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IBIKKG.

