

## Analisis Kritis dan Pengembangan Algoritma K-Nearest Neighbor (KNN): Sebuah Tinjauan Literatur

Raja Sakti Arief Daulay<sup>1)\*</sup>

<sup>1)</sup>STKIP Pangeran Antasari, Indonesia

<sup>1)</sup>[rajasaktidaulay@gmail.com](mailto:rajasaktidaulay@gmail.com)



\*Raja Sakti Arief Daulay

### Histori Artikel:

Submit: 2024-12-05

Diterima: 2024-12-13

Dipublikasikan: 2024-12-13

### Kata Kunci:

K-Nearest Neighbors; Machine Learning; Peningkatan Algoritma; Metode Klasifikasi; Data Mining (minimal 5 kata)

### ABSTRAK

K-Nearest Neighbors (KNN) adalah sebuah algoritma klasifikasi populer yang sering digunakan dalam *machine learning*. KNN mengklasifikasikan objek berdasarkan data pembelajaran yang memiliki jarak terdekat dengan objek tertentu. Algoritma *Nearest Neighbor* hanya mempertimbangkan satu data pembelajaran terdekat sebagai basis klasifikasi. Walaupun KNN sederhana tapi kekuatan algoritma ini terletak pada kemampuannya menghasilkan prediksi yang cukup akurat dalam berbagai kasus aplikasi. Namun, kelemahan utama algoritma ini adalah sensitifitasnya terhadap parameter seperti jumlah tetangga ( $k$ ) dan metrik jarak yang digunakan. Seiring waktu, banyak peneliti telah melakukan berbagai pengembangan dan improvisasi terhadap KNN untuk meningkatkan akurasi hasil klasifikasinya. Berbagai pendekatan telah diterapkan, seperti optimasi pemilihan parameter, pengembangan metode pengukuran jarak, hingga penerapan algoritma hibrida dengan teknik *machine learning* lainnya. Improvisasi yang dikembangkan mencakup penyesuaian adaptif pada nilai  $k$ , penggunaan metrik jarak yang lebih kompleks, serta integrasi metode optimisasi seperti pembobotan jarak dan teknik *ensemble*. Penelitian-penelitian tersebut menunjukkan hasil yang signifikan dalam meningkatkan performa KNN, baik dalam hal efisiensi maupun akurasi. Paper ini bertujuan untuk mengkaji berbagai jenis pengembangan yang telah dilakukan pada algoritma KNN, menganalisis keunggulan dan kelemahannya, serta mengidentifikasi faktor-faktor yang mempengaruhi kinerja algoritma ini. Dengan memahami berbagai inovasi yang telah dilakukan, diharapkan dapat memberikan wawasan mendalam terkait potensi peningkatan lebih lanjut pada algoritma KNN. Selain itu, penelitian ini juga membahas penerapan KNN dalam berbagai domain, seperti pengenalan pola, prediksi, dan pengolahan data skala besar. Hasil kajian ini diharapkan dapat menjadi acuan bagi pengembang algoritma untuk memanfaatkan keunggulan KNN secara lebih optimal serta membuka peluang penelitian baru dalam bidang klasifikasi berbasis jarak.

Jurnal Pendidikan Sains dan Komputer is licensed under a Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0).

### LATAR BELAKANG

Algoritma KNN sudah banyak dipakai untuk penelitian yang berhubungan dengan machine learning. Berbagai macam peran digunakan untuk metode pengklasifikasian yang dilakukan oleh KNN. Oleh karena itu banyak juga penelitian yang sudah memodifikasi atau meningkatkan kinerja dari KNN tersebut. Dimulai dengan meningkatkan akurasi pada pengklasifikasian, menggabungkan KNN dengan algoritma. Klasifikasi merupakan masalah yang penting dalam mengolah big data, data science, dan machine learning. KNN adalah salah satu algoritma tertua, paling sederhana, dan juga akurat untuk klasifikasi pola dan model regresi. KNN diajukan pada tahun 1951 oleh Evelyn dan Hodges dan kemudian dimodifikasi oleh Cover dan Hart, KNN telah diidentifikasi sebagai salah satu dari sepuluh metode teratas dalam data mining (L. Wang 2019). Akibatnya, KNN telah dipelajari selama beberapa dekade terakhir dan diterapkan secara luas di berbagai bidang. Dengan demikian, KNN terdiri dari pengklasifikasi dasar dalam banyak masalah klasifikasi pola seperti pengenalan pola, kategorisasi teks, model peringkat, pengenalan objek, dan aplikasi

pengenalan peristiwa (Mehta et al. 2021). KNN adalah algoritma nonparametric. Nonparametrik berarti tidak ada parameter atau jumlah parameter tetap terlepas dari ukuran data. Sebaliknya, parameter akan ditentukan oleh ukuran kumpulan data pelatihan, meskipun tidak ada asumsi yang perlu dibuat untuk distribusi data yang mendasarinya. Dengan demikian, KNN bisa menjadi pilihan terbaik untuk setiap studi klasifikasi yang melibatkan sedikit atau tanpa pengetahuan sebelumnya tentang distribusi data. Selain itu, KNN adalah salah satu metode pembelajaran yang paling malas (Tamrakar and Syed Ibrahim 2021). Ini berarti penyimpanan semua data pelatihan akan menunggu sampai data uji dihasilkan, tanpa harus membuat model pembelajaran. Kelambatan bukan satu-satunya masalah yang terkait dengan pengklasifikasi KNN, selain memilih masalah K-neighbors terbaik, memilih ukuran jarak/kesamaan terbaik adalah masalah penting, hal ini karena kinerja pengklasifikasi KNN bergantung pada jarak/ukuran kesamaan yang digunakan (Abu Alfeilat et al. 2019; Uddin et al. 2022).

Kelemahan pada algoritma KNN ini banyak dijadikan sebagai bahan para peneliti untuk berupaya agar meningkatkan kinerja dari KNN. Sehingga banyak jenis dari peningkatan ataupun modifikasi dari algoritma KNN. Banyak juga KNN digabungkan dengan teknik machine learning lainnya sehingga dapat mengurangi dari kelemahan KNN tersebut. Oleh karena itu penelitian akan berfokus mengenalkan beberapa modifikasi dan peningkatan dari algoritma KNN, sehingga dengan harapan dapat menjadi pedoman untuk penelitian di masa yang akan datang.

#### STUDI LITERATUR

Beberapa penelitian sudah banyak meneliti tentang peningkatan algoritma KNN dan pemodifikasian dari algoritma KNN. Setiap penelitian memiliki variasi studi yang berbeda-beda mulai dari data distribusi, tipe data, dan menggunakan jumlah jarak dan ukuran kesamaan yang berbeda. Pengembangan KNN ini jelas banyak dilakukan karena algoritma ini cukup sederhana tetapi efektif untuk masalah klasifikasi dan regresi. Proses pengembangan algoritma KNN secara umum dapat dilakukan dengan pengumpulan data, pemilihan fitur, penentuan nilai  $k$ , perhitungan jarak, penentuan kelas atau nilai dan evaluasi model. Dengan mengikuti langkah sederhana itu maka pengembangan KNN dapat dilakukan dan di aplikasikan untuk jenis masalah klasifikasi.

Seperti pada penelitian ini (An, Xu, and Shen 2019) yang melakukan peningkatan KNN untuk klasifikasi metode pembelajaran, dimana pada proses klasifikasi yang menggunakan KNN tradisional diganti dengan diusulkan algoritma KNN yang berdasarkan skema density tailoring. Pada penelitian lainnya (B. Wang et al. 2020) disini mengusulkan metode *weighted* KNN (WKNN) yang dimana metode ini adalah sebuah peningkatan KNN yang digunakan untuk mencapai akurasi pemosisian tingkat ruangan. Algoritma ini menggunakan pengelompokan propagasi afinitas untuk mengelompokkan *Reference Point* (RP) terdekat sesuai dengan jarak posisi mereka satu sama lain dan *outlier* disaring untuk memisahkan *subset* dengan jumlah RP yang lebih besar yang dimana pada hasilnya dengan meningkatkan algoritma ini membuat keakuratan yang baik.

#### METODE

Penelitian ini merupakan kajian pustaka dari beberapa artikel yang terkait algoritma KNN. Peninjauan dilakukan dari beberapa upaya penelitian terbaru yang memanfaatkan algoritma KNN. Setelah itu kajian ini berasal dari beberapa literasi dan mencakup upaya peningkatan serta modifikasi pada algoritma KNN yang dibagi ke dalam setiap judul penelitian.

Proses pengumpulan data yang digunakan untuk memeriksa beberapa literature sangat berguna untuk mencari dan memperoleh sumber kajian berdasarkan penelitian relevan sebelumnya. Teori-teori yang

mendukung, data dan informasi sebagai referensi dalam dokumentasi. Pada penelitian ini di tabel 1 akan dijelaskan lebih detail tentang peningkatan algoritma yang sudah pernah dilakukan sebelumnya dan perbandingan dari jenis-jenis algoritma KNN.

### HASIL

Hasil dari ulasan menunjukkan beberapa penelitian menggunakan algoritma KNN yang ditingkatkan untuk memecahkan masalah sesuai dengan kebutuhan pada bidang penelitian masing-masing. Penjelasan mengenai peningkatan algoritma KNN dan mengenai pemecahan masalahnya secara detail ditunjukkan pada tabel 1.

Tabel 1  
Perbandingan jenis pengembangan dengan algoritma KNN

Judul Penelitian	Keterangan
<b>Effects of Distance Measure Choice on K-Nearest Neighbor Classifier Performance: A Review</b>	Menganalisis kinerja pengklasifikasi KNN menggunakan 11 ukuran jarak. Termasuk <i>Euclidean Distance</i> (ED), <i>Manhattan Distance</i> (MD), <i>Minkowski Distance</i> , <i>Chebyshev Distance</i> , <i>Cosine Distance</i> (CosD), <i>Correlation Distance</i> (CorD), <i>Hamming Distance</i> (HamD), <i>Jaccard Distance</i> (JacD), <i>Standardized Euclidean Distance</i> , dan <i>Spearman Distance</i> . Penelitian ini telah diterapkan pada delapan set data sintetik biner dengan berbagai jenis distribusi yang dihasilkan menggunakan MATLAB. Mereka membagi setiap set data menjadi 70% untuk set pelatihan dan 30% untuk set pengujian. Hasilnya menunjukkan bahwa ukuran MD, <i>Minkowski Distance</i> , <i>Chebyshev Distance</i> , ED, <i>Mahalanobis Distance</i> , dan <i>Standardized Euclidean Distance</i> mencapai hasil akurasi yang sama dan mengungguli jarak yang diuji lainnya. Kumpulan data dinormalisasi sebelum melakukan percobaan. Untuk mengevaluasi kinerja KNN, pengukuran akurasi, sensitivitas, dan spesifitas dihitung untuk setiap jarak. Hasil yang dilaporkan menunjukkan bahwa penggunaan MD mengungguli jarak yang diuji lainnya, dengan tingkat akurasi 97,8%, tingkat sensitivitas 96,76%, dan tingkat spesifisitas 98,35%. Tidak ada metric jarak optimal yang dapat digunakan untuk semua jenis kumpulan data, karena hasilnya menunjukkan bahwa setiap kumpulan data menyerupai metrik jarak tertentu. Performa (diukur dengan akurasi, presisi, dan recall) dari KNN hanya turun 20% sedangkan noise level mencapai 90%, ini berlaku untuk semua distance yang digunakan. Ini berarti bahwa pengklasifikasi KNN yang menggunakan salah satu dari 10 distance teratas menoleransi noise hingga tingkat tertentu (Abu Alfeilat et al. 2019).
<b>A new locally adaptive k-nearest neighbor algorithm based on discrimination class</b>	Metode pemilihan <i>nearest neighbour</i> dari algoritma KNN tidak cukup handal. Salah satu alasannya adalah karena mengabaikan pengaruh distribusi spasial dari kueri dan <i>instance</i> pelatihan pada klasifikasi. Oleh karena itu konsep baru <i>nearest centroid neighbor</i> (NCN) diusulkan, yang mempertimbangkan distribusi contoh pelatihan di lingkungan kueri. Algoritma <i>dependent nearest neighbor</i> (dNN) yang mempertimbangkan tidak hanya kesamaan yang ditentukan oleh jarak tetapi juga ketergantungan yang ditentukan dengan <i>angle</i> ketika memilih <i>nearest neighbor</i> yang bergantung dari kueri. Alasan lain adalah bahwa metode pemilihan <i>nearest neighbor</i> dari algoritma KNN hanya menggunakan kesamaan sepihak dilihat dari sudut pandang kueri. Untuk meningkatkan kesamaan antara kueri dan <i>nearest neighbor</i> , informasi <i>neighbor</i> pada contoh pelatihan telah diperhitungkan dalam banyak pekerjaan. Algoritma KNN hanya menggunakan <i>neighbor</i> kelas mayoritas di <i>k-neighborhood</i> dari kueri untuk menentukan hasil klasifikasi kueri dan sepenuhnya mengabaikan informasi termasuk <i>nearest neighbor</i> di kelas lain yang dapat menyebabkan hasil klasifikasi yang buruk ketika jumlah <i>nearest neighbor</i> di kelas yang berbeda tidak jauh berbeda. Selain itu, algoritma KNN menggunakan nilai k tetap dan tunggal untuk semua kueri, tetapi k ini tidak

	dapat optimal untuk kueri di lokasi spasial yang berbeda(Pan, Wang, and Pan 2020).
<b>Study and Observation of the Variation of Accuracies of KNN, SVM, LMNN, ENN Algorithms on Eleven Different Datasets from UCI Machine Learning Repository</b>	Algoritma KNN dan SVM mirip seperti dalam kedua kasus di area tertentu yang dianggap menemukan seberapa besar probabilitas. KNN yang ditingkatkan menunjukkan akurasi yang lebih baik pada lokasi teks web dalam kelompok daripada algoritma KNN otomatis. Manfaat utama SVM adalah kemampuannya menangani sejumlah besar kumpulan data yang gagal dilakukan KNN dalam beberapa kasus. Disisi lain kelemahannya adalah <i>cost</i> dan kerumitannya yang tinggi. Untuk mengatasi kekurangan di KNN, metode lain bernama <i>Extended Nearest Neighbor</i> (ENN) diusulkan dan kinerjanya diamati yang menyiratkan bahwa ENN menunjukkan kinerja yang lebih baik daripada KNN tradisional(Na et al. 2021). Di ENN selain mempertimbangkan sampel uji data terdekat juga dihitung dalam data terdekat yang memberikan akurasi yang lebih baik. Pada keempat algoritma <i>machine learning</i> yaitu KNN, ENN, SVM, dan LMNN telah menunjukkan dampak kuat pada klasifikasi data di berbagai sektor. Meskipun saat ini KNN dan SVM digantikan oleh ENN dan LMNN namun karena lebih sempurna dan akurat, peran vital KNN dan SVM tidak dapat diabaikan(Khan et al. 2018).
<b>DDoS Attack Detection Method Based on Improved KNN with the Degree of DDoS Attack in Software-Defined Networks</b>	Serangan <i>Distributed Denial of Service</i> (DDoS) telah mengganggu ketersediaan jaringan secara berulang beberapa dekade dan masih belum ada mekanisme pertahanan yang efektif untuk melawannya. Dalam penelitian ini mereka mengusulkan dua metode untuk mendeteksi serangan DDoS di SDN. Salah satu metode mengadopsi tingkat serangan DDoS untuk mengidentifikasi serangan DDoS. Metode lain menggunakan algoritma <i>K-Nearest Neighbors</i> (KNN) yang ditingkatkan berdasarkan <i>Machine Learning</i> untuk menemukan serangan DDoS. Menggabungkan algoritma klasifikasi <i>Support Vector Machine</i> (SVM) untuk membangun model serangan DDoS(Dong and Sarem 2020). Dalam metode mereka, enam nilai fitur diperkenalkan. Hasil eksperimen mereka menunjukkan tingkat alarm palsu yang rendah untuk trafik TCP dan UDP, tetapi tingkat alarm palsu trafik ICMP tinggi. Mengusulkan mekanisme menggunakan algoritma klasifikasi <i>Cognitive-Inspired Computing</i> (CIC) dan SVM untuk mendeteksi serangan DDoS. Sementara, itu akurasi pendeteksiannya masih perlu ditingkatkan lagi. Berdasarkan dari konsep penelitian ini, algoritma pendeteksi yang disebut <i>Algorithm based in the Degree of Attack</i> (DDADA) diusulkan. Selain itu untuk lebih meningkatkan efisiensi deteksi, algoritma deteksi lain yang disebut algoritma DDoS <i>Detection Algorithm based on Machine Learning</i> (DDAML) diperkenalkan untuk mengidentifikasi serangan DDoS. Hasil eksperimen menunjukkan bahwa algoritma yang diusulkan dapat mengidentifikasi serangan DDoS dengan lebih baik dan telah mencapai tingkat deteksi yang lebih tinggi dibandingkan dengan solusi yang ada. Hasil eksperimen menunjukkan bahwa algoritma DDAML dapat mengungguli algoritma lain pada pengukuran kinerja yang berbeda.
<b>Classification Method of Teaching Resources Based on Improved KNN Algorithm</b>	Pemodelan kinerja personel manufaktur berdasarkan algoritma KNN yang ditingkatkan. Sejak itu, banyak ahli dan peneliti telah mencapai hasil penelitian tertentu di bidang teknologi ini, seperti ilmuwan intelijen terkenal Spark dan Salton. Sejak 1980-an, teknik rekayasa pengetahuan tradisional telah diterapkan di bidang ini. Menurut pengetahuan yang diberikan oleh para ahli, aturan dibentuk dan pengklasifikasi dibuat secara manual. Ini adalah klasifikasi yang baik di beberapa korpora. Namun, dalam menghadapi kumpulan data skala besar, metodenya terbatas. Proses <i>preprocessing teks</i> ditingkatkan dengan strategi yang diusulkan menggunakan algoritma KNN. Tujuan pada masalah algoritma KNN tradisional dalam klasifikasi sumber daya pengajaran di sekolah dasar dan menengah maka diusulkan algoritma KNN yang baik

	<p>berdasarkan skema <i>density tailoring</i>. Sampel daerah dengan kepadatan tinggi di ruang sampe dipotong sebelum algoritma KNN dijalankan. Masalah kesalahan klasifikasi disebabkan oleh distribusi kepadatan spasial sampel yang tidak merata diselesaikan. Pada saat yang sama, kompleksitas waktu klasifikasi KNN berkurang. Parameter yang sesuai K dan <i>input</i> ditentukan oleh percobaan perbandingan dan efektivitas algoritma yang ditingkatkan telah diverifikasi. Hasil penelitian menunjukkan bahwa metode klasifikasi sumber data ajar SD dan SMP berbasis algoritma KNN layak dan efektif (An, Xu, and Shen 2019).</p>
<b>Fast density peak clustering for large scale data based on kNN</b>	<p><i>Vantage-point tree</i> atau <i>VP tree</i> sangat mirip dengan <i>K-D tree</i> yang memisahkan data dalam ruang metrik dengan memilih posisi dalam ruang (titik pandang) dan membagi titik data menjadi dua partisi: yang lebih dekat ke <i>vantage point</i> daripada <i>threshold</i> dan yang tidak. Dengan berulang kali menerapkan prosedur ini untuk mempartisi data menjadi set yang lebih kecil dan lebih kecil, struktur data <i>tree</i> dibuat dimana <i>neighbor</i> di <i>tree</i> cenderung menjadi <i>neighbor</i> di ruang. FLANN adalah <i>library</i> untuk melakukan <i>fast approximate nearest neighbor</i> di ruang dimensi tinggi. FLANN berisi kumpulan algoritma untuk pencarian <i>neighbor</i> terdekat dan sistem untuk secara otomatis memilih algoritma terbaik dan parameter optimal tergantung pada kumpulan data. Menggunakan <i>cover tree</i> untuk menerima KNN untuk setiap titik dan KNN-<i>density</i> diusulkan untuk menggantikan kepadatan asli yang didefinisikan dalam <i>Dpeak</i>, yang menghasilkan peningkatan besar untuk perhitungan kepadatan (Y. Chen et al. 2020).</p>
<b>Cost-Sensitive KNN Classification</b>	<p>Metode KNN untuk mengklasifikasikan <i>big data</i> (Ma and Chi 2022). KNN pertama melakukan <i>k-means clustering</i> untuk memisahkan seluruh <i>dataset</i> menjadi beberapa bagian. Kemudian setiap <i>subset</i> diklasifikasikan dengan metode KNN. <i>kTree</i> dan <i>k*Tree</i> untuk menggunakan jumlah <i>neighbor</i> terdekat yang berbeda untuk klasifikasi KNN. Metode <i>kTree</i> membutuhkan <i>cost</i> operasional yang lebih sedikit tetapi mencapai akurasi klasifikasi yang serupa dibandingkan dengan metode KNN yang menetapkan nilai K yang berbeda untuk sampel uji yang berbeda. Metode <i>k*Tree</i> adalah perpanjangan dari <i>kTree</i>. Yaitu mempercepat tahap pengujiannya dengan menyimpan informasi sampel pelatihan di <i>leafnode ktree</i>, seperti sampel pelatihan yang terletak di <i>leafnode</i> KNN mereka dan <i>neighbor</i> terdekat dari KNN. Membuat KNN hanya menggunakan subset sampel pelatihan di <i>leafnode</i>. Paper ini menyajikan dua pendekatan yaitu pengklasifikasi <i>direct-cs-knn</i> dan pengklasifikasi <i>distance-cs-knn</i> yang bertujuan untuk membuat klasifikasi KNN <i>sensitive</i> terhadap <i>cost</i> sehingga dapat meminimalkan <i>cost</i> kesalahan klasifikasi. Untuk efisiensi beberapa metode yang berguna termasuk <i>smoothing</i> pengaturan K dengan biaya minimum, pemilihan fitur CS dan CS <i>stacking</i> secara signifikan digabungkan ke pengklasifikasi CS-KNN (S. Zhang 2020).</p>
<b>Machine learning methods for cyber security intrusion detection: Datasets and comparative study</b>	<p>Dalam studi ini (Kilincer, Ertam, and Sengur 2021) peneliti menggunakan metode <i>deep learning</i> (<i>recurrent neural network</i>) RNN, <i>deep neural networks</i> (DNN), <i>restricted Boltzmann machines</i> (RBM), <i>deep belief networks</i> (DBN), <i>convoluted neural networks</i> (CNN), <i>deep Boltzmann machines</i> (DBM), dan <i>deep autoencoders</i> (DA) telah mengimplementasikannya pada dataset CSE-CIC-IDS2018 dan Bot-lot. Kemudian keberhasilan klasifikasi <i>deep learning</i> dan waktu klasifikasi dari kumpulan data ini dibandingkan pada metode <i>deep learning</i> diperiksa dan dalam pengertian ini 35 set data deteksi serangan yang digunakan dalam literature dibagi ke dalam kategori. Dataset CSE-CIC-IDS-2017, karena dataset yang ada tidak memenuhi kebutuhan deteksi intrusi saat ini. Pengujian <i>enviromtent</i> yang terdiri dari penyerang dan korban jaringan telah disiapkan untuk membuat kumpulan data. Di pengujian <i>enviromtent</i> serangan seperti <i>brute force</i>, <i>heartbleed attack</i>, <i>botnet</i>, DOS, DDoS, <i>web attack</i>, <i>infiltration attack</i> diatur. Selain itu kinerja sistem dievaluasi</p>



<b>Internet Digital Economy Development Forecast Based on Artificial Intelligence and SVM-KNN Network Detection</b>	<p>menggunakan metode <i>machine learning</i>. Dengan perkembangan teknologi pintar, internet digunakan di setiap bidang kehidupan sehari-hari. Dengan meluasnya penggunaan internet, jenis serangan berkembang dari hari ke hari. Untuk mencegah serangan, serangan ini harus dideteksi terlebih dahulu. Dalam hal ini sistem IDS telah dikembangkan untuk mendeteksi lalu lintas serangan dan set data IDS telah dibuat untuk mensimulasikan jenis serangan (Al-E'mari et al. 2022; Khraisat et al. 2019).</p> <p>Beberapa penelitian mengusulkan model pengelompokan KNN berdasarkan puncak kepadatan yang dapat mendeteksi serangan lebih efektif dan memperkenalkan kepadatan ke KNN. Proses KNN tidak memerlukan banyak parameter dan proses <i>iterative</i> didasarkan pada densitas. Literatur mengusulkan model deteksi intrusi Vervis baru dimana metode klasifikasi data mining baru digabungkan yaitu sebagai alternatif untuk jaringan Vervis menggunakan algoritma. Berdasarkan penelitian literatur terkait algoritma klasifikasi SVM-KNN dan algoritma klasifikasi KNN, sesuai dengan karakteristik klasifikasi dari algoritma klasifikasi KNN saat ini dan kombinasi keunggulan dari algoritma KNN berbobot harmonik dan algoritma <i>support vector machine</i>, kombinasi dari algoritma SVM dan KNN akan digabungkan untuk membangun sebuah algoritma KNN berbobot harmonik berbasis SVM untuk meningkatkan kinerja klasifikasi dari kumpulan data (Fu, Zhou, and Mei 2022).</p>
<b>Comparison and analysis of logistic regression, Naive Bayes and KNN machine learning algorithms for credit card fraud detection</b>	<p>Meneliti dan memeriksa penyajian algoritma pengklasifikasi <i>Decision Tree</i>, <i>Random Forest</i>, SVM dan <i>Logistic Regression</i>. Metode yang digunakan pada informasi mentah dan pra-ditangani. Dari investigasi yang telah dilakukan didapatkan hasil bahwa Regresi Logistik memiliki ketepatan 97,7% sedangkan SVM menunjukkan ketepatan 97,5% dan <i>decision tree</i> menunjukkan ketepatan 95,5% namun hasil terbaik diperoleh dengan <i>random forest</i> dengan presisi tepat 98,6%. Hasil yang diperoleh karena itu alasan bahwa <i>random forest</i> menunjukkan akurasi yang paling tepat dan tinggi dari 98,6% dalam masalah deteksi penipuan kartu kredit dengan dataset yang diberikan oleh ULB. Memeriksa berbagai pengklasifikasi pembelajaran mesin yang dilatih pada kumpulan data publik untuk menganalisis korelasi faktor-faktor tertentu dengan penipuan. Metrik yang lebih baik digunakan untuk menentukan tingkat negatif palsu dan kinerja pengambilan sampel acak diukur untuk menangani ketidakseimbangan kelas dari kumpulan data (Ittoo, Meenakshi, and Singh 2021). SVM berkinerja lebih baik untuk mendeteksi penipuan kartu kredit dalam kondisi yang realistis. Perbandingan antara model algoritma pembelajaran mendalam dan regresi dilakukan untuk menentukan algoritma dan kombinasi faktor mana yang memberikan metode paling akurat untuk mengklasifikasikan transaksi kartu kredit sebagai penipuan atau non-penipuan. Algoritma terbaik untuk analisis kumpulan data dengan rasio transaksi penipuan dan non-penipuan mendekati 1:1 adalah <i>Random Forest Classifier</i>, dengan asumsi distribusi penipuan ke tidak penipuan dari set pengujian dan pelatihan adalah sama (Abidin et al. 2020; Hamsa et al. 2021; Zhou et al. 2021). <i>Logistic Regression</i> (LR) menunjukkan kinerja yang optimal untuk semua proporsi data dibandingkan dengan <i>Naïve Bayes</i> (NB) dan <i>K-Nearest Neighbor</i> (KNN). LR berhasil mendapatkan akurasi yang lebih tinggi dibandingkan dengan <i>Naïve Bayes</i> dan KNN. LR menunjukkan akurasi maksimum 95%, NB menunjukkan 91% dan KNN 75%. Teknik LR juga menunjukkan <i>Sensitivity</i>, <i>Specificity</i>, <i>Precision</i> dan <i>F-Measure</i> yang lebih baik dibandingkan dengan teknik NB dan KNN. Juga telah diamati bahwa menjadi teknik yang diawasi (LR dan <i>Naïve Bayes</i>) menunjukkan hasil yang lebih baik dalam setiap kasus dibandingkan dengan teknik KNN yang tidak diawasi.</p>

<b>An Enhanced Human Speech Emotion Recognition Using Hybrid of PRNN and KNN</b>	<p>Pengklasifikasi KNN berbasis jarak rata-rata komprehensif melalui biaya rata-rata multi-komprehensif serta bentangan rata-rata komprehensif terstruktur yang didasarkan pada perwakilan rata-rata komprehensif diusulkan pada tahun 2019. Dalam metode yang diusulkan, multi-lokal vektor rata-rata dari sampel kueri yang diberikan di setiap kelas dihitung dengan mengadopsi <i>neighbor K-Nearest</i> spesifik kelasnya. Dengan diperoleh <math>k</math> vektor rata-rata lokal per kelas, <math>k</math> ruang rata-rata umum ditentukan dan digunakan dalam merancang bentangan rata-rata komprehensif <i>nested</i> yang pasti. Teknik <i>ensemble</i> baru untuk algoritma KNN juga diusulkan pada tahun 2019. Algoritma <i>ensemble</i> berbasis gangguan multimodal yang disebut <i>Reduced Random Subspacebases Bagging</i> (RRSB) diusulkan yang menghasilkan pengklasifikasi komponen yang akurat namun beragam untuk meningkatkan kinerja klasifikasi <i>ensemble</i>. Nilai <math>k</math>, sistem RRSB yang diusulkan tampaknya lebih kuat daripada teknik lainnya. Dalam teknik yang diusulkan ini, pengenalan emosi ucapan yang ditingkatkan dilakukan pada enam emosi dasar marah, senang, sedih, netral, terkejut, dan takut. Sebagai metodologi penelitian lanjutan, pra-pemrosesan dilakukan menggunakan algoritma <i>Pattern Recognition Neural Network</i> (PRNN) dan KNN sedangkan ekstraksi ciri dilakukan dengan menggunakan struktur bertingkat yang terdiri dari <i>Mel Frequency Cepstral Coefficient</i> (MFCC) dan <i>Gray Level Co-Occurrence Matrix</i> (GLCM). Hasil yang diperoleh dibandingkan untuk akurasi, tingkat presisi dan <math>f</math> Measure dengan algoritma standar seperti <i>Gaussian Mixture Model</i> (GMM) dan <i>Hidden Markov Model</i> (HMM) dan diakui sebagai <i>output</i> yang lebih baik daripada algoritma standar (Umamaheswari and Akila 2019).</p>
<b>A Novel Weighted KNN Algorithm Based on RSS Similarity and Position Distance for Wi-Fi Fingerprint Positioning</b>	<p>Mengusulkan metode <i>Weighted KNN</i> (WKNN) untuk menetapkan bobot yang berbeda dengan mendefinisikan koefisien korelasi antara <i>Access Point</i> dan mencapai akurasi pemosisian tingkat ruangan. Mengusulkan algoritma WKNN yang difilter cluster. Algoritma pengelompokan propagasi afinitas untuk mengelompokkan <i>Reference Point</i> (RP) terdekat sesuai dengan jarak posisi mereka satu sama lain, dan <i>outlier</i> disaring untuk memesan subset dengan jumlah RP yang lebih besar. RP dengan metode ini adalah tidak didasarkan pada jarak posisi antara RP dan <i>Test Point</i> (TP). Oleh karena itu, beberapa RP yang dekat dengan TP dapat dibuang sebagai <i>outlier</i>, yang mengakibatkan kesalahan penentuan posisi yang besar. Dengan menggabungkan jarak <i>weighted Euclidean</i> dengan informasi posisi yang diketahui dari RP, perkiraan posisi dirancang dan digunakan untuk meningkatkan algoritma. Keakuratan posisi rata-rata dari algoritma yang diusulkan mengungguli <i>Euclidean-WKNN</i> sebesar 45,28%, 38,41%, dan 20,60%, mengungguli algoritma berbasis SVM masing-masing sebesar 40,36%, 37,94% dan 33,74% di tiga database (B. Wang et al. 2020).</p>
<b>A novel ensemble method for k-nearest neighbour</b>	<p>Banyak makalah penelitian telah menyelidiki ansambel KNN dengan tujuan meningkatkan kinerjanya. Seperti, penelitian ini (Y. Zhang et al. 2019) menerapkan beberapa metrik jarak untuk menghasilkan anggota ansambel yang beragam, dimana metrik jarak diperlukan sebagai gangguan parameter pembelajaran. Penelitian Ishii (Ishii et al. 2005) menggabungkan KNN dengan menggunakan <i>Genetic Algorithm</i> (GA) untuk menimbang fungsi jarak yang berbeda. Beberapa subruang acak digunakan untuk mendapat pengklasifikasi KNN komponen diselidiki oleh Ho yang melatih setiap KNN pada subset atribut acak daripada seluruh ruang atribut. Dalam penelitian ini diusulkan algoritma baru multimodal <i>perturbation-based ensemble, reduced random subspace-based bagging</i> (RRSB) yang menghasilkan pengklasifikasi komponen yang akurat namun beragam untuk meningkatkan kinerja klasifikasi <i>ensemble</i>. Hasil eksperimen dari beberapa set data UCI menunjukkan bahwa metode yang diusulkan dapat meningkatkan kinerja klasifikasi dalam banyak</p>

	<p>kasus. Dibandingkan dengan metode lain, RRSB kuat dengan nilai <math>k</math> yang berbeda.</p>
<b>Explaining and Improving Model Behavior with <math>k</math> Nearest Neighbor Representations</b>	<p>Telah diperkenalkan KNN <i>language</i> model (KNN-LM) memperluas model bahasa pra-terlatih yang ada dengan menginterpolasi distribusi kata berikutnya dengan model KNN. Kombinasi KNN-LM merupakan <i>met-learner</i> yang dapat secara efektif disetel untuk menghafal dan mengambil pola <i>long-tail</i> yang langka (Rajani et al. 2020). Pekerjaan ini mengadaptasi KNN-LM ke klasifikasi teks dengan menggunakan contoh input sebagai konteks dengan cara yang mirip dengan penerapan <i>deep</i>-KNN melalui representasi <i>neural network</i> untuk klasifikasi gambar. Penelitian ini menemukan bahwa KNN untuk setiap contoh pengujian dapat memberikan informasi yang berguna tentang bagaimana model membuat keputusan klasifikasinya, untuk salah mengklasifikasi contoh. Dari analisis juga mengamati bahwa KNN dari contoh uji yang salah klasifikasi sering kali menunjukkan contoh yang salah label, yang memberikan pendekatan ini aplikasi untuk memperbaiki contoh yang salah label dalam set pelatihan. Dengan menemukan <i>nearest neighbor</i> yang paling umum di seluruh set pengujian, mereka juga dapat mengidentifikasi subset dari contoh pelatihan yang sangat berpengaruh dan mendapatkan interpretasi tingkat korpus untuk kinerja model. Terakhir, yaitu memeriksa utilitas untuk mundur ke klasifikasi berbasis KNN ketika kepercayaan model rendah. Analisis batas keputusan yang dipelajari oleh KNN atas representasi tersembunyi menunjukkan bahwa KNN mempelajari keputusan yang <i>fine-grained</i> batas yang dapat membantu membuatnya lebih kuat terhadap perubahan kecil dalam teks yang menyebabkan label kebenaran dasar, tetapi bukan prediksi model terbalik.</p>
<b>Voice Disorder Identification by using Hilbert-Huang Transform (HHT) and K Nearest Neighbor (KNN)</b>	<p>Saat menentukan metode klasifikasi, KNN adalah klasifikasi sampel menurut kategori sampel terdekat. Dibandingkan dengan metode lain, KNN lebih cocok untuk partisi kumpulan sampel di mana domain kelas berpotongan atau tumpang tindih (L. Chen et al. 2021). Kompleksitas pelatihannya lebih rendah daripada SVM. Selain itu, dibandingkan dengan algoritma seperti <i>Naïve Bayes</i>, KNN tidak memiliki asumsi tentang data, akurasi tinggi, dan tidak sensitive terhadap titik abnormal. KNN digunakan di banyak bidang. Xinyu Li (Li et al. 2018) menambahkan algoritma KNN ke sistem <i>machine vision</i> untuk mewujudkan pemantauan <i>real-time</i> dari kesalahan kecil. Deteksi otomatis gangguan suara memiliki signifikansi klinis yang sangat penting untuk mendiagnosis disfonía (Tartarisco et al. 2021). Makalah ini menyajikan metode klasifikasi otomatis untuk gangguan suara dan suara normal yang berbasis HHT dan KNN. Hasil eksperimen menunjukkan bahwa metode ini memiliki performansi yang baik, akurasi 93,3%, presisi 93%, <i>recall rate</i> 95%, <i>F1-score</i> 94%, dan AUC 97,6%. Hasil eksperimen memverifikasi validitas dan reliabilitas metode ekstraksi ciri. Eksperimen menunjukkan bahwa model KNN tidak hanya mendapatkan hasil klasifikasi yang baik, tetapi juga meningkatkan kinerja generalisasi <i>classifier</i>.</p>

## PEMBAHASAN

Dengan adanya peningkatan dari algoritma KNN diharapkan tidak menutup kemungkinan hanya dengan data pengetahuan yang terbatas namun dapat mengembangkan lagi varian-varian dari KNN. Tidak semua teknik bisa dikembangkan untuk KNN tetapi didalam tabel diatas adalah sebuah bukti bahwa algoritma yang ada dapat dikembangkan sehingga menjadi keterbaruan pengetahuan. Algoritma KNN pun terus dikembangkan oleh beberapa peneliti.



---

### KESIMPULAN

Algoritma KNN pada dasarnya adalah algoritma *supervised learning* yang digunakan untuk melakukan klasifikasi atau regresi berdasarkan tetangga terdekat dari suatu data yang baru. Dari tabel diatas maka dapat diteliti ada beberapa teknik yang dapat digunakan untuk meningkatkan kinerja algoritma KNN, di antaranya:

1. Standarisasi data: Sebelum memasukkan data ke dalam algoritma KNN, penting untuk melakukan standarisasi data. Hal ini dilakukan untuk memastikan bahwa setiap variabel memiliki skala yang sama dan memperbaiki distribusi data yang tidak normal
2. Penggunaan Bobot: Memberikan bobot setiap tetangga yang berbeda-beda berdasarkan jarak antara data dan titik yang ingin diprediksi. Semakin dekat jaraknya, maka bobotnya semakin besar. Hal ini dapat meningkatkan akurasi prediksi.
3. Pemilihan fitur: Memilih fitur yang paling relevan dan memiliki pengaruh besar pada hasil prediksi. Hal ini dapat membantu mengurangi dimensi data dan mempercepat waktu komputasi.
4. *Cross-validation*: Teknik ini dapat membantu menghindari *overfitting* dan *underfitting* pada model KNN. *Cross-validation* dapat membantu mengevaluasi performa model pada data yang tidak terlihat sebelumnya.
5. Penggunaan metrik jarak yang sesuai: Memilih metrik jarak yang sesuai untuk masalah yang ingin diselesaikan. Beberapa contoh metrik jarak yang umum digunakan adalah *Euclidean*, *Manhattan*, dan *Cosine distance*.

Dengan menerapkan teknik-teknik ini, kita dapat meningkatkan kinerja algoritma KNN dan memperoleh prediksi yang lebih akurat.

### REFERENSI

- Abidin, Nur Haziqah Zainal, Muhammad Akmal Remli, Noorlin Mohd Ali, Danakorn Nincarean Eh Phon, Nooraini Yusoff, Hasyiya Karimah Adli, and Abdelsalam H. Busalim. 2020. "Improving Intelligent Personality Prediction Using Myers-Briggs Type Indicator and Random Forest Classifier." *International Journal of Advanced Computer Science and Applications*. doi:10.14569/IJACSA.2020.0111125.
- Abu Alfeilat, Haneen Arafat, Ahmad B.A. Hassanat, Omar Lasassmeh, Ahmad S. Tarawneh, Mahmoud Bashir Alhasanat, Hamzeh S. Eyal Salman, and V. B.Surya Prasath. 2019. "Effects of Distance Measure Choice on K-Nearest Neighbor Classifier Performance: A Review." *Big Data* 7(4): 221–48. doi:10.1089/big.2018.0175.
- Al-E'mari, Salam, Mohammed Anbar, Yousef Sanjalawe, Selvakumar Manickam, and Iznan Hasbullah. 2022. "Intrusion Detection Systems Using Blockchain Technology: A Review, Issues and Challenges." *Computer Systems Science and Engineering*. doi:10.32604/CSSE.2022.017941.
- An, Yingbo, Meiling Xu, and Chen Shen. 2019. "Classification Method of Teaching Resources Based on Improved KNN Algorithm." *International Journal of Emerging Technologies in Learning* 14(4): 73–88. doi:10.3991/ijet.v14.i04.10131.
- Chen, Lili, Chaoyu Wang, Junjiang Chen, Zejun Xiang, and Xue Hu. 2021. "Voice Disorder Identification

- 
- by Using Hilbert-Huang Transform (HHT) and K Nearest Neighbor (KNN).” *Journal of Voice* 35(6): 932.e1-932.e11. doi:10.1016/j.jvoice.2020.03.009.
- Chen, Yewang, Xiaoliang Hu, Wentao Fan, Lianlian Shen, Zheng Zhang, Xin Liu, Jixiang Du, et al. 2020. “Fast Density Peak Clustering for Large Scale Data Based on KNN.” *Knowledge-Based Systems* 187(xxxx). doi:10.1016/j.knosys.2019.06.032.
- Dong, Shi, and Mudar Sarem. 2020. “DDoS Attack Detection Method Based on Improved KNN with the Degree of DDoS Attack in Software-Defined Networks.” *IEEE Access* 8: 5039–48. doi:10.1109/ACCESS.2019.2963077.
- Fu, Jianru, Xu Zhou, and Guoping Mei. 2022. “Internet Digital Economy Development Forecast Based on Artificial Intelligence and SVM-KNN Network Detection.” *Computational Intelligence and Neuroscience* 2022. doi:10.1155/2022/5792694.
- Hamsa, Shibani, Youssef Iraqi, Ismail Shahin, and Naoufel Werghi. 2021. “An Enhanced Emotion Recognition Algorithm Using Pitch Correlogram, Deep Sparse Matrix Representation and Random Forest Classifier.” *IEEE Access*. doi:10.1109/ACCESS.2021.3086062.
- Ishii, Naohiro, Eisuke Tsuchiya, Yongguang Bao, and Nobuhiko Yamaguchi. 2005. “Combining Classification Improvements by Ensemble Processing.” In *Proceedings - Third ACIS International Conference on Software Engineering Research, Management and Applications, SERA 2005*, doi:10.1109/SERA.2005.30.
- Itoo, Fayaz, Meenakshi, and Satwinder Singh. 2021. “Comparison and Analysis of Logistic Regression, Naïve Bayes and KNN Machine Learning Algorithms for Credit Card Fraud Detection.” *International Journal of Information Technology (Singapore)* 13(4): 1503–11. doi:10.1007/s41870-020-00430-y.
- Khan, Mohammad Mahmudur Rahman, Rezoana Bente Arif, Abu Bakr Siddique, and Mahjabin Rahman Oishe. 2018. “Study and Observation of the Variation of Accuracies of KNN, SVM, LMNN, ENN Algorithms on Eleven Different Datasets from UCI Machine Learning Repository.” *4th International Conference on Electrical Engineering and Information and Communication Technology, iCEEiCT 2018*: 124–29. doi:10.1109/CEEICT.2018.8628041.
- Khraisat, Ansam, Iqbal Gondal, Peter Vamplew, and Joarder Kamruzzaman. 2019. “Survey of Intrusion Detection Systems: Techniques, Datasets and Challenges.” *Cybersecurity*. doi:10.1186/s42400-019-0038-7.
- Kilincer, Ilhan Firat, Fatih Ertam, and Abdulkadir Sengur. 2021. “Machine Learning Methods for Cyber Security Intrusion Detection: Datasets and Comparative Study.” *Computer Networks* 188(October 2020): 107840. doi:10.1016/j.comnet.2021.107840.
- Li, Xinyu, Tiezhu Qiao, Yusong Pang, Haitao Zhang, and Gaowei Yan. 2018. “A New Machine Vision Real-Time Detection System for Liquid Impurities Based on Dynamic Morphological Characteristic Analysis and Machine Learning.” *Measurement: Journal of the International Measurement Confederation*. doi:10.1016/j.measurement.2018.04.015.
- Ma, Chen, and Yuhong Chi. 2022. “KNN Normalized Optimization and Platform Tuning Based on Hadoop.” *IEEE Access*. doi:10.1109/ACCESS.2022.3195872.

- Mehta, Ankush, Deepam Goyal, Anurag Choudhary, B. S. Pabla, and Safya Belghith. 2021. "Machine Learning-Based Fault Diagnosis of Self-Aligning Bearings for Rotating Machinery Using Infrared Thermography." *Mathematical Problems in Engineering*. doi:10.1155/2021/9947300.
- Na, Junying, Zhiping Wang, Siqi Lv, and Zhaohui Xu. 2021. "An Extended K Nearest Neighbors-Based Classifier for Epilepsy Diagnosis." *IEEE Access*. doi:10.1109/ACCESS.2021.3081767.
- Pan, Zhibin, Yikun Wang, and Yiwei Pan. 2020. "A New Locally Adaptive K-Nearest Neighbor Algorithm Based on Discrimination Class." *Knowledge-Based Systems* 204: 106185. doi:10.1016/j.knsys.2020.106185.
- Rajani, Nazneen Fatema, Ben Krause, Wengpeng Yin, Tong Niu, Richard Socher, and Caiming Xiong. 2020. "Explaining and Improving Model Behavior with k Nearest Neighbor Representations." <http://arxiv.org/abs/2010.09030>.
- Tamrakar, Preeti, and S. P. Syed Ibrahim. 2021. "Lazy Learning Associative Classification with WkNN and DWkNN Algorithm." *ITM Web of Conferences*. doi:10.1051/itmconf/20213701023.
- Tartarisco, Gennaro, Roberta Bruschetta, Susanna Summa, Liliana Ruta, Martina Favetta, Mario Busa, Alberto Romano, et al. 2021. "Artificial Intelligence for Dysarthria Assessment in Children with Ataxia: A Hierarchical Approach." *IEEE Access*. doi:10.1109/ACCESS.2021.3135078.
- Uddin, Shahadat, Ibtisham Haque, Haohui Lu, Mohammad Ali Moni, and Ergun Gide. 2022. "Comparative Performance Analysis of K-Nearest Neighbour (KNN) Algorithm and Its Different Variants for Disease Prediction." *Scientific Reports*. doi:10.1038/s41598-022-10358-x.
- Umamaheswari, J., and A. Akila. 2019. "An Enhanced Human Speech Emotion Recognition Using Hybrid of PRNN and KNN." *Proceedings of the International Conference on Machine Learning, Big Data, Cloud and Parallel Computing: Trends, Perspectives and Prospects, COMITCon 2019*: 177–83. doi:10.1109/COMITCon.2019.8862221.
- Wang, Boyuan, Xingli Gan, Xuelin Liu, Baoguo Yu, Ruicai Jia, Lu Huang, and Haonan Jia. 2020. "A Novel Weighted KNN Algorithm Based on RSS Similarity and Position Distance for Wi-Fi Fingerprint Positioning." *IEEE Access* 8: 30591–602. doi:10.1109/ACCESS.2020.2973212.
- Wang, Lishan. 2019. "Research and Implementation of Machine Learning Classifier Based on KNN." In *IOP Conference Series: Materials Science and Engineering*, doi:10.1088/1757-899X/677/5/052038.
- Zhang, Shichao. 2020. "Cost-Sensitive KNN Classification." *Neurocomputing* 391: 234–42. doi:10.1016/j.neucom.2018.11.101.
- Zhang, Youqiang, Guo Cao, Bisheng Wang, and Xuesong Li. 2019. "A Novel Ensemble Method for K-Nearest Neighbor." *Pattern Recognition* 85: 13–25. doi:10.1016/j.patcog.2018.08.003.
- Zhou, Tao, Libin Chen, Jing Guo, Mengmeng Zhang, Yanrui Zhang, Shanbo Cao, Feng Lou, and Haijun Wang. 2021. "MSIFinder: A Python Package for Detecting MSI Status Using Random Forest Classifier." *BMC Bioinformatics*. doi:10.1186/s12859-021-03986-z.