



ANALISA KANKER PARU PARU DENGAN MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBOR

LUNG CANCER ANALYSIS USING K-NEAREST NEIGHBOR ALGORITHM

Teguh Abdi Mangun¹⁾, Odi Nurdiawan^{*2)}, Ade Irma Purnamasari^{*3)}

¹⁾Program Studi Teknik Informatika, STIMIK IKMI Cirebon

Jl. Perjuangan No.10B, Karyamulya, Kesambi, Kota Cirebon, Jawa Barat 45131, Indonesia.

Abstract – Lung cancer is one of the deadliest types of cancer in the world. According to data from the World Health Organization (WHO), lung cancer is the type of cancer that causes the most deaths worldwide. Therefore, the development of accurate and effective classification methods for lung cancer is essential in efforts to improve early detection and appropriate treatment. Stages of lung cancer analysis research using the k-Nearest Neighbor (k-NN) algorithm. The initial stage in this research is collecting data related to lung cancer. After the data is collected, the next step is preprocessing the data to clean it. Selecting the k value, namely the number of nearest neighbors, is a crucial step in lung cancer analysis. Once the k value is selected, a k-NN model is trained using the training data to learn the relationship between attributes and lung cancer status. The results of this research are that the accuracy obtained is 80.40%.

Keywords – Lungs, KNN Algorithm, Cigarettes.

Abstrak – Kanker paru adalah salah satu jenis kanker yang paling mematikan di dunia. Menurut data dari World Health Organization (WHO), kanker paru merupakan jenis kanker yang paling banyak menyebabkan kematian di seluruh dunia. Karena itu, pengembangan metode klasifikasi yang akurat dan efektif untuk kanker paru sangat penting dalam upaya untuk meningkatkan deteksi dini dan pengobatan yang tepat. Tahapan penelitian analisis kanker paru-paru menggunakan algoritma k-Nearest Neighbor (k-NN). Tahap awal dalam penelitian ini adalah mengumpulkan data terkait kanker paru-paru. Setelah data terkumpul, langkah berikutnya adalah melakukan preprocessing data untuk membersihkan. Pemilihan nilai k, yaitu jumlah tetangga terdekat, merupakan langkah krusial dalam analisis kanker paru-paru. Setelah nilai k terpilih, model K-NN dilatih menggunakan data pelatihan untuk mempelajari hubungan antara atribut dan status kanker paru-paru. Hasil pada penelitian ini yaitu hasil akurasi yang didapat yaitu sebesar 80.40%.

Kata kunci – Paru, Algoritma KNN, Kebiasaan merokok.

I. PENDAHULUAN

Kanker paru adalah salah satu jenis kanker yang paling mematikan di dunia. Menurut data dari World Health Organization (WHO), kanker paru merupakan jenis kanker yang paling banyak menyebabkan kematian di seluruh dunia. Karena itu, pengembangan metode klasifikasi yang akurat dan efektif untuk kanker paru sangat penting dalam upaya untuk meningkatkan deteksi dini dan pengobatan yang tepat. Menurut data dari Globocan 2020, diperkirakan terdapat sekitar 57.500 kasus baru kanker paru di Indonesia pada tahun 2020 dan sekitar 45.000 kematian akibat kanker paru. Angka ini menunjukkan bahwa kanker paru merupakan salah satu jenis kanker yang paling mematikan di Indonesia. Selain itu, menurut data dari Kementerian Kesehatan Republik Indonesia, kanker paru juga menjadi jenis kanker yang paling banyak diderita oleh laki-laki di Indonesia dengan persentase sekitar 17,8% dari seluruh jenis kanker yang terjadi pada laki-laki, dan pada perempuan sekitar 9,1%. Tingginya angka kejadian kanker paru di Indonesia menunjukkan bahwa upaya pencegahan dan deteksi dini kanker paru sangat penting untuk dilakukan. Oleh karena itu, edukasi dan kampanye tentang bahaya meKebiasaan merokok dan faktor risiko lainnya perlu dilakukan secara masif untuk menurunkan angka kejadian kanker paru di Indonesia. (Derajad Wijaya & Dwiasnati, 2020)

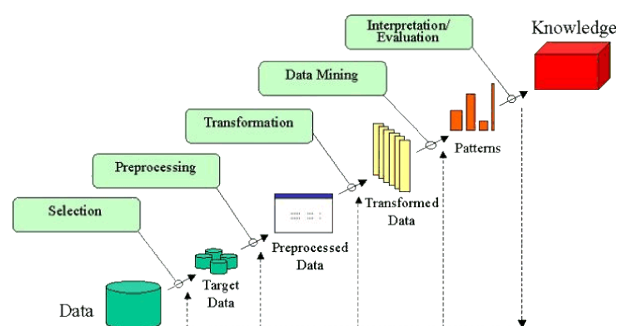
Salah satu metode yang saat ini sedang berkembang untuk klasifikasi kanker paru adalah menggunakan algoritma K-Nearest Neighbor. K-Nearest Neighbor adalah sebuah model matematika yang mirip dengan cara kerja otak manusia, yang terdiri dari lapisan-lapisan neuron dan dapat dilatih untuk melakukan klasifikasi pada data yang kompleks. Penelitian tentang implementasi algoritma K-Nearest Neighbor untuk klasifikasi kanker paru sangat penting karena dapat memberikan kontribusi dalam meningkatkan akurasi dan efektivitas deteksi dini kanker paru. Selain itu, penelitian ini juga dapat membantu dokter dalam membuat keputusan yang lebih baik dalam pengobatan kanker paru, sehingga dapat meningkatkan tingkat kesembuhan dan kualitas hidup pasien. (Nainel et al., 2020)

Kebaruan dalam penelitian ini data yang bersumber dari kaggle.com memiliki data yang lebih banyak, kemudian metode yang digunakan yaitu neural network yang memiliki pembelajaran yang baik, dengan parameter yang digunakan yaitu shuffle dengan tujuan dapat meningkatkan akurasi.

K-Nearest Neighbors (KNN) adalah salah satu algoritma klasifikasi yang digunakan dalam machine learning. Algoritma ini bekerja berdasarkan prinsip bahwa objek-objek yang serupa cenderung terletak di sekitar satu sama lain dalam ruang fitur. Penting untuk diingat bahwa KNN adalah salah satu dari banyak algoritma klasifikasi yang tersedia, dan tidak selalu menjadi pilihan terbaik untuk setiap kasus. Pemilihan algoritma yang tepat harus mempertimbangkan karakteristik khusus dari data dan masalah yang dihadapi.

III. METODE PENELITIAN

Metode penelitian sebagai alur penelitian yang dapat digunakan dalam hasil dan pembahasan sehingga penelitian ini terarah, adapun alur penelitian sebagai berikut :



Gambar 1. Alur Penelitian

A. Pengumpulan Data

Melakukan identifikasi sumber data yang relevan seperti basis data medis, penelitian sebelumnya, atau data dari lembaga kesehatan. Pilih atribut atau fitur yang dianggap penting untuk analisis, seperti usia, jenis kelamin, riwayat meKebiasaan merokok, hasil tes medis, dll.

B. Preprocessing Data:

Memersihkan dan persiapan data, termasuk menangani nilai-nilai yang hilang, mengatasi outlier, dan melakukan normalisasi jika diperlukan. Tentukan kelas atau label target (menderita kanker paru-paru atau tidak). (Sitanggang et al., 2022)

C. Transformation

Transformasi atau transformasi data adalah proses mengubah atau memanipulasi data asli untuk memperbaiki kualitas atau mempersiapkannya sebelum dianalisis. Dalam analisis kanker paru-paru menggunakan algoritma K-Nearest Neighbors (KNN), transformasi data dapat menjadi langkah penting untuk meningkatkan kualitas dan kinerja model (Munandar & Munir, 2022)

D. Datamining Algoritma K-Nearest Neighbor

Bagi data menjadi set pelatihan (training set) dan set pengujian (testing set) untuk melatih dan menguji model K-NN. Implementasikan algoritma K-Nearest Neighbor untuk melatih model klasifikasi berdasarkan data yang telah diproses. (Novitasari et al., 2019)

E. Interpretasi Hasil:

Analisis hasil dari model K-NN untuk memahami atribut yang paling berpengaruh dalam klasifikasi. Identifikasi faktor risiko yang signifikan. (Novitasari et al., 2019)

III. HASIL DAN PEMBAHASAN

Hasil Penelitian Analisa Kanker Paru Paru Dengan Menggunakan Algoritma K-Nearest Neighbor.

A. Data Selection

Penelitian ini menggunakan data yang diambil dari Kaggle.com, dengan fokus pada kriteria kanker paru-paru. Data diambil pada bulan September 2023 dan dapat diakses melalui tautan berikut: <https://www.kaggle.com/datasets/thedevastator/cancer-patients-and-air-pollution-a-new-link>.

Populasi penelitian terdiri dari pasien penderita kanker paru-paru pada tahun 2023, dengan jumlah populasi sekitar 1001 data pasien. Parameter yang digunakan dalam penelitian ini mencakup informasi tentang usia, jenis kelamin, paparan polusi udara, konsumsi alkohol, alergi debu, bahaya pekerjaan, risiko genetik, penyakit paru-paru kronis, pola makan seimbang, kebiasaan berat badan, kebiasaan meKebiasaan merokok, paparan asap Kebiasaan merokok pasif, gejala nyeri dada, batuk darah, kelelahan, penurunan berat badan, kesulitan bernapas, mengi, kesulitan menelan, kuku jari tabuh, dan mendengkur. Kemudian data yang digunakan dalam penelitian ini sebanyak 1001 data pasien yang akan dilakukan klasifikasi penderita kanker paru paru berdasarkan kriteria di tabel 4.1 kemudian hasilnya menjadi high dan low.

B. Preproceesing

Data harus diproses untuk membuatnya lebih mudah diolah oleh jaringan saraf. Ini termasuk pembersihan data, normalisasi, dan pemilihan fitur yang relevan. Dalam penelitian ini ada beberapa data yang

tidak lengkap sehingga peneliti melakukan pembersihan dengan menghapus data tersebut.

C. Data Transformation

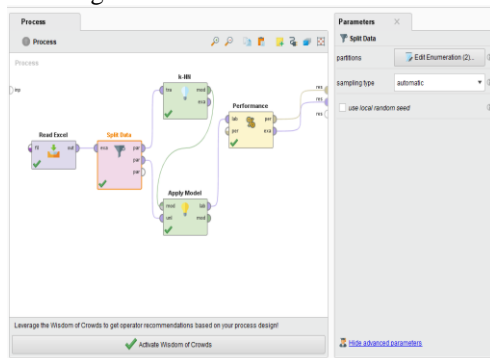
Data transformation adalah proses mengubah format, struktur, atau nilai dari data mentah untuk membuatnya lebih sesuai untuk analisis atau penggunaan tertentu. Ini adalah salah satu tahap penting dalam persiapan data sebelum dilakukan analisis atau pembuatan model



Gambar2. Transformation

D. Data Mining Algoritma K-Nearest Neighbor

Model Algoritma *K-Nearest Neighbor* dapat dimodelkan sebagai berikut :

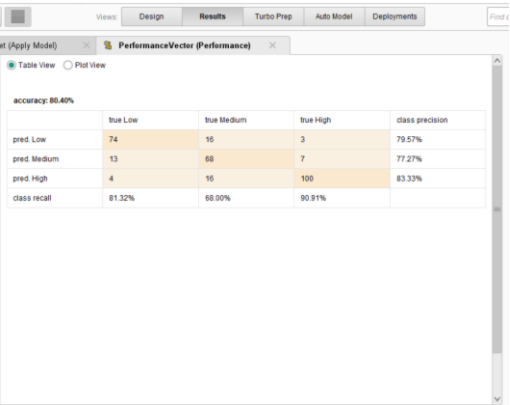


Gambar 3. Model Algoritma K-Nearest Neighbor

Berdasarkan model algoritma *K-Nearest Neighbor* yang tertuang dalam gambar 4.2 sehingga membutuhkan Operator excel, Operator Split Data, Operator K-NN, Operator Apply Model dan Operator Performance.

E. Evaluation

Evaluation pada tahap ini yaitu menampilkan hasil akurasi sebagaimana dalam gambar berikut ini :



	True Low	True Medium	True High	Class precision
pred. Low	74	16	3	79.57%
pred. Medium	13	68	7	77.27%
pred. High	4	16	109	83.33%
Class recall	81.32%	68.00%	90.91%	

Gambar 4. Hasil Akurasi confusion matrix

IV. KESIMPULAN

Penelitian bertujuan untuk mengevaluasi kinerja algoritma K-Nearest Neighbors (k-NN) dalam menganalisis data terkait kanker paru-paru dengan tujuan memprediksi kemungkinan keberadaan penyakit. Hasil akurasi yang didapat yaitu sebesar 80.40%. Dalam konteks penelitian ini, kinerja algoritma K-Nearest Neighbors dievaluasi dengan fokus pada prediksi penyakit kanker paru-paru, dan hasilnya diukur dengan menggunakan confusion matrix untuk mengukur akurasi klasifikasi. Penyesuaian parameter k adalah langkah penting dalam meningkatkan kinerja model dan memberikan hasil yang lebih baik dalam deteksi penyakit.

Hasil penelitian mengindikasikan bahwa peningkatan akurasi dan sensitivitas dalam analisis kanker paru-paru menggunakan algoritma k-NN dapat dicapai dengan menyesuaikan parameter k. Hal ini dapat dicapai dengan memilih nilai k yang tepat, mempertimbangkan metrik jarak yang sesuai, dan melibatkan teknik preprocessing data yang cermat.

UCAPAN TERIMA KASIH

Kami ingin mengungkapkan rasa terima kasih kami kepada seluruh staf pustaka yang dengan sabar membantu kami dalam menemukan sumber-sumber yang relevan untuk penelitian ini. Bantuan Anda sangat berarti bagi perkembangan penelitian kami..

DAFTAR PUSTAKA

- [1] Adriani, a. (2022). Prediksi kandidat protein target antikanker derivat kumarin asal tanaman sanrego (*lunasia amara blanco*). in *biology and education journal* (vol. 2, issue 2). <https://prediction.charite.de>
- [2] Derajad Wijaya, H., & Wiasnati, S. (2020). Implementasi data mining dengan algoritma naïve bayes pada penjualan obat. *jurnal informatika*, 7(1). <http://ejournal.bsi.ac.id/ejurnal/index.php/ji>



- [3] Fardian Anshori, i., & Riana, d. (2021). Prediksi harapan hidup pasien kanker paru-paru pasca operasi bedah thoraks menggunakan boosted neural network dan smote. *jurnal infomedia: teknik informatika, multimedia dan janringan*, 6(1). <http://archive.ics.uci.edu/ml/datasets>.
- [4] Munandar, t. A., & Munir, a. Q. (2022). Implementasi k-nearest neighbor untuk prototype sistem pakar identifikasi dini penyakit jantung k-nearest neighbor for prototype expert system for early identification of heart disease. *Jurnal teknologi informasi*.
- [5] Nainel, y. L., Buulolo, e., & Lubis, i. (2020). Penerapan data mining untuk estimasi penjualan obat berdasarkan pengaruh brand image dengan algoritma expectation maximization (studi kasus: pt. pyridam farma tbk). *jurikom (jurnal riset komputer)*, 7(2), 214. <https://doi.org/10.30865/jurikom.v7i2.2097>
- [6] Novitasari, h. B., Hadiano, n., Sfenrianto, Rahmawati, a., Prasetyo, r., Miharja, j., & Gata, w. (2019). K-nearest neighbor analysis to predict the accuracy of product delivery using administration of raw material model in the cosmetic industry (pt cedefindo). *journal of physics: conference series*, 1367(1). <https://doi.org/10.1088/1742-6596/1367/1/012008>
- [7] Priyanti, e. (2021). Bianglala informatika penerapan algoritma neural network untuk klasifikasi kanker paru. *jurnal bianglala informatika*, 9(1), 2021.
- [8] Purba, w., Wardani, s., Lumbantoruan, d. f., Celia, f., silalahi, i., & leo edison, t. (2023). Optimization of lung cancer classification method using eda-based machine learning. *journal of information systems and computer science prima*, 6(2).
- [9] Sari, l., Romadloni, a., Listyaningrum, r., studi, p., informatika, t., cilacap, p. n., & soetomo, j. (2023). Penerapan data mining dalam analisis prediksi kanker paru menggunakan algoritma random forest. 14(01). <https://doi.org/10.35970/infotekmesin.v14i1.1751>
- [10] Sitanggang, d., Nicholas, n., Wilson, v., Sinaga, a. r. a., & Simanjuntak, a. d. (2022). implementasi data mining untuk memprediksi penyakit jantung menggunakan metode k-nearest neighbor dan logistic regression. *jurnal teknik informasi dan komputer (tekinkom)*, 5(2), 493. <https://doi.org/10.37600/tekinkom.v5i2.698>