



Optimasi K-Nearest Neighbor dengan Grid Search CV pada Prediksi Kanker Paru-Paru

Satya Tegar Kusuma¹, Theopilus Bayu Sasongko²

satyategar@students.amikom.ac.id, theopilus.27@amikom.ac.id

Universitas Amikom Yogyakarta

Informasi Artikel

Diterima : 9 Jul 2023
Direview : 31 Jul 2023
Disetujui : 29 Ags 2023

Kata Kunci

Kanker paru-paru, *Grid search cv*, *K-nearest neighbor*, Optimasi

Abstrak

Kanker paru-paru adalah salah satu kanker paling mematikan di seluruh. Salah satu penyebab kematian pada penderita kanker paru-paru adalah tidak ada sistem untuk memprediksi kanker paru-paru secara optimal apakah pasien menderita kanker paru-paru atau tidak. Oleh karena itu, penelitian ini bertujuan untuk melakukan optimasi nilai K pada algoritma *k-nearest neighbor* (KNN) menggunakan metode *Grid Search CV*. Algoritma KNN dipilih karena pada berbagai penelitian memiliki tingkat akurasi yang lebih baik dibandingkan dengan algoritma *supervised learning* lainnya. Data yang digunakan pada penelitian ini bersumber dari data publik yang ada di *Kaggle*. Berdasarkan penelitian dan pembahasan mengenai optimasi nilai K pada algoritma KNN menggunakan metode *grid search cv* didapatkan nilai K paling optimal yaitu 3 dengan tingkat akurasi 96%. Oleh karena itu, nilai $k=3$ sangat baik diterapkan pada algoritma KNN untuk memprediksi kanker paru-paru karena memiliki akurasi yang tinggi.

Keywords

Lung Cancer, *Grid search*, *K-nearest neighbor*, *Optimization*

Abstrak

Lung cancer is one of the deadliest cancers around. One of the causes of death in lung cancer patients is that there is no system to predict lung cancer optimally whether the patient has lung cancer or not. Therefore, this study aims to optimize the value of K in the k-nearest neighbor (KNN) algorithm using the cv grid search method. The KNN algorithm was chosen because in various studies it has a better level of accuracy compared to other supervised learning algorithms. The data used in this study comes from public data on Kaggle. Based on research and discussion on optimizing the K value in the KNN algorithm using the cv grid search method, the optimal K value is 3 with an accuracy rate of 96%. Therefore, the value of $K = 3$ is very well applied to the KNN algorithm to predict lung cancer because it has high accuracy.

A. Pendahuluan

Kanker paru-paru adalah salah satu kanker yang paling mematikan di seluruh dunia. Pada tahun 2020 terjadi 2,2 juta kasus kanker paru-paru di seluruh dunia [1]. Di Indonesia terdapat 396.914 kasus di tahun 2020, dengan pembagian kanker paru-paru 34.783 (8,8%) [2]. Banyaknya kasus terjadi disebabkan karena kurangnya sistem untuk memprediksi apakah seseorang menderita kanker paru-paru atau tidak. Untuk memprediksi kanker paru-paru membutuhkan data yang kompleks dan ukuran sampelnya yang besar. Oleh karena itu, diperlukan sebuah metode yang efektif, efisien, dan akurat untuk memprediksi kanker paru-paru.

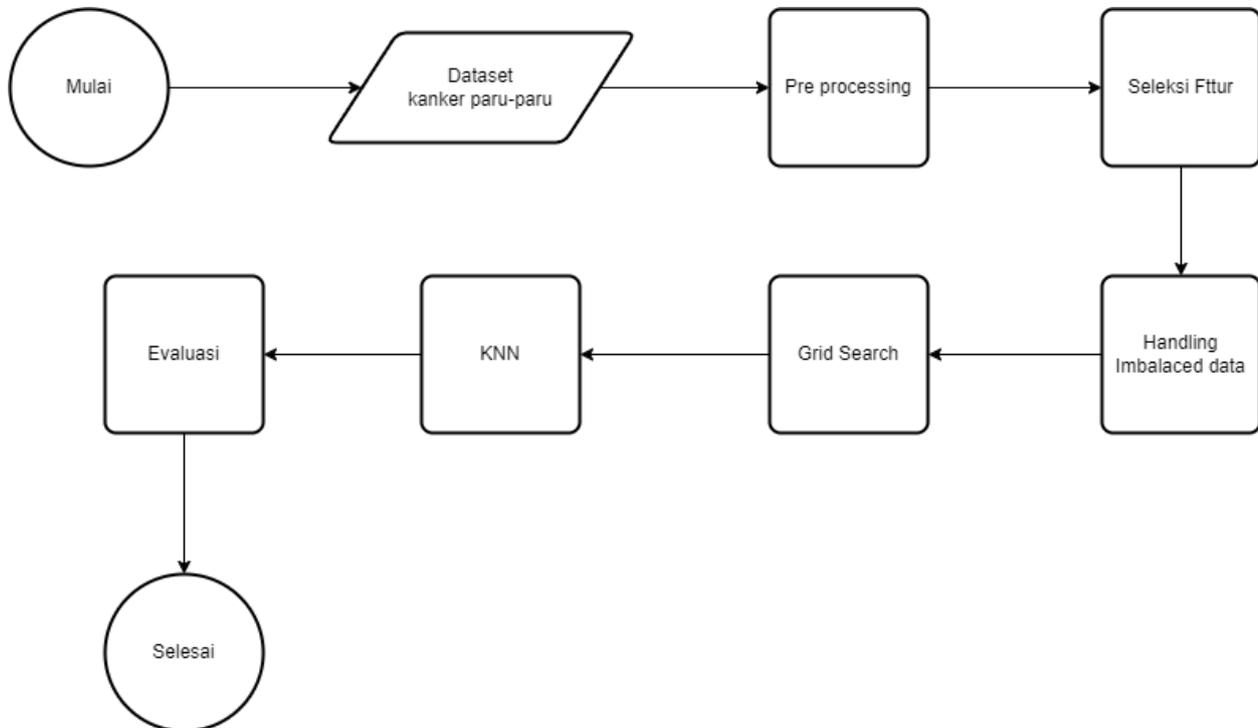
Salah satu metode yang sering digunakan dalam klasifikasi adalah *k-nearest neighbor* (KNN). KNN adalah salah satu algoritma sederhana yang memanfaatkan informasi dari tetangga terdekat untuk mengklasifikasikan suatu sampel. Namun, performa KNN sangat bergantung pada pemilihan parameter seperti nilai K dan fungsi jarak yang digunakan. Selain itu, dalam klasifikasi kanker paru-paru keakuratan KNN dapat dipengaruhi oleh jumlah fitur yang digunakan dan bagaimana fitur-fitur tersebut dipilih. Dalam penelitian tentang perbandingan akurasi algoritma *k-nearest neighbor* dan *logistic regression* untuk klasifikasi penyakit diabetes, KNN mendapat akurasi sebesar 85,06%, sedangkan *logistic regression* mendapatkan akurasi 77,92% [3]. Penelitian lainnya tentang perbandingan perbandingan algoritma KNN, *decision tree*, dan *random forest* pada data *imbalanced class* untuk klasifikasi promosi karyawan mendapat hasil akurasi 85,57% yang merupakan akurasi terbaik diantara algoritma lainnya [4]. Dari penelitian-penelitian tersebut KNN adalah algoritma dengan akurasi terbaik dibandingkan dengan algoritma lainnya.

Setelah algoritma dengan akurasi terbaik ditemukan, muncul masalah baru yaitu algoritma KNN, nilai K belum dioptimasi atau dengan kata lain penentuan nilai K masih manual. Oleh sebab itu, perlu dilakukan sebuah optimasi nilai secara otomatis tanpa harus menguji satu persatu secara manual. Dalam penelitian ini akan menggunakan *grid search cv*, sebuah metode untuk menemukan optimasi nilai yang dalam penggunaannya dikombinasikan dengan *cross validation*. Metode ini pernah digunakan dalam penelitian optimasi *support vector machine* (SVM) menggunakan *grid search* untuk mengidentifikasi citra biji kopi robusta berdasarkan *circularity* dan *eccentricity* yang mendapatkan akurasi 94% [5]. Penelitian yang menggunakan *grid search* juga terbukti meningkatkan nilai akurasi pada penelitian *machine learning models for prognosticating dementia* [6]. Pada penelitian untuk klasifikasi sarana akses air yang membandingkan algoritma KNN dengan *naïve bayes* juga mendapatkan hasil yang serupa yaitu KNN memiliki akurasi yang lebih baik [7].

Oleh karena masalah-masalah tersebut, penelitian ini bertujuan untuk melakukan optimasi nilai K pada algoritma *k-nearest neighbor* menggunakan *grid search cv* untuk meningkatkan performa dalam memprediksi kanker paru-paru. Penelitian ini dilakukan menggunakan data publik tentang kanker paru-paru. Penelitian ini diharapkan dapat memberikan kontribusi penting bagi pengembangan metode klasifikasi paru-paru yang efektif, efisien, dan akurat.

B. Metode Penelitian

Penelitian ini menggunakan *grid search* untuk optimasi nilai K pada metode *k-nearest neighbor* untuk prediksi kanker paru-paru. Penggambaran tahapan ini ditunjukkan pada Gambar 1.



Gambar 1. Tahapan Penelitian

1. Data Kanker Paru-Paru

Dataset kanker paru-paru diambil dari data publik yang ada di *kaggle* yang berisi 309 data dengan 16 fitur [8]. Pada dataset tersebut dilakukan klasifikasi data kanker paru-paru dengan target YES (menderita) dan NO (tidak menderita), persebaran datanya dapat dilihat pada Tabel 1.

Tabel 1. Data Kanker Paru-Paru

| No | Kanker Paru-Paru | Jumlah |
|----|------------------|--------|
| 1 | YES | 270 |
| 2 | NO | 39 |

2. Pre-Processing

Pada tahap ini, tipe data pada dataset disamakan menjadi numerik (int). Proses penyamaan tipe data menggunakan teknik *encoding* dengan mengubah tipe data pada kolom GENDER dan LUNG_CANCER.

Tabel 2. Data Kanker Sebelum Encoding

| NO | GENDER | LUNG_CANCER | TIPE DATA |
|----|--------|-------------|-----------|
| 1 | M | YES | Objek |
| 2 | F | NO | Objek |

Sebelum dilakukan *encoding* kolom GENDER berisi M dan F, sedangkan kolom LUNG_CANCER berisi YES dan NO seperti pada Tabel 2. Kedua kolom tersebut memiliki tipe data berupa objek yang akan bermasalah pada saat dilakukan korelasi antar tabel.

Tabel 3. Data Kanker Setelah Encoding

| No | Gender | Lung_Cancer | Tipe Data |
|----|--------|-------------|-----------|
| 1 | 1 | 1 | Int |
| 2 | 0 | 0 | Int |

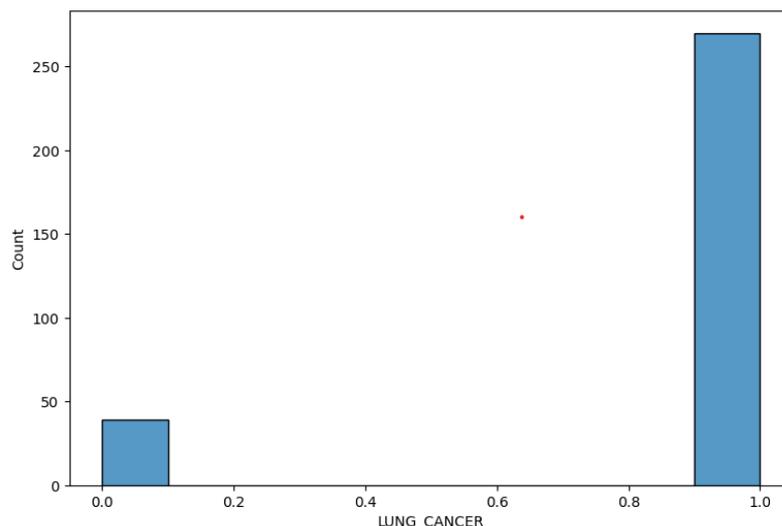
Oleh karena itu, harus dilakukan *encoding* untuk mengubah tipe data kedua kolom menjadi int (kategorikal) seperti pada Tabel 3 sehingga pada saat dilakukan korelasi antar tabel tidak terjadi masalah.

3. Seleksi Fitur

Seleksi fitur merupakan proses pemilihan atribut yang sesuai dari sekumpulan atribut yang banyak sehingga pada akhirnya algoritma yang digunakan untuk memproses data lebih cepat karena telah dilakukan pengurangan atribut [9]. Dalam proses *data mining*, seleksi fitur sangat penting dan mempunyai pengaruh yang sangat signifikan. Selain untuk mempercepat algoritma yang dipilih untuk memproses data, seleksi fitur juga digunakan untuk meningkatkan akurasi dari algoritma tersebut. Hal tersebut dikarenakan fitur-fitur yang digunakan tentu sangat berpengaruh dengan target *mining* sehingga akan menghasilkan akurasi yang lebih baik juga. Pada penelitian ini, jumlah Fitur pada data sebelum dilakukan seleksi adalah 16, kemudian dilakukan korelasi antar tabel untuk diambil nilai korelasi lebih dari 0,1 terhadap target. Dari tahap korelasi dihasilkan fitur hanya berjumlah 12.

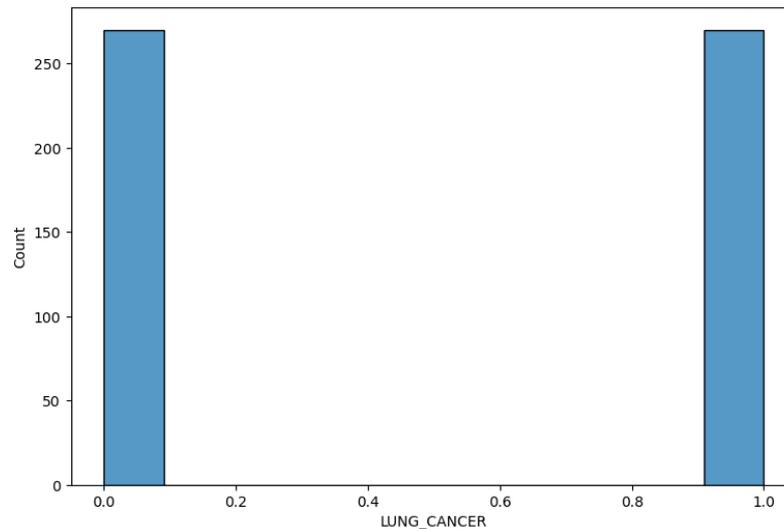
4. Handling Imbalance Data

SMOTE adalah metode yang menerapkan *oversampling* dalam mengatasi data yang kurang seimbang dengan mempertimbangkan efek pada wilayah keputusan dalam ruang lingkup [10]. Jika dibandingkan dengan metode *undersampling*, SMOTE lebih baik karena tidak ada pengurangan data.



Gambar 2. Sebelum SMOTE

Pada Gambar 2 terlihat data yang tidak seimbang antara 0 dan 1. 0 hanya memiliki data kurang dari 50, sedangkan 1 memiliki data lebih dari 250.

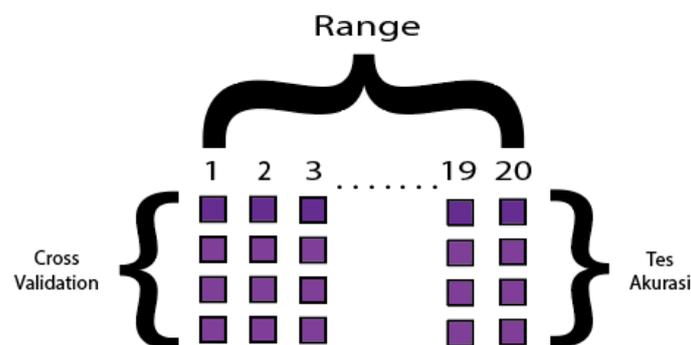


Gambar 3. Setelah SMOTE

Pada Gambar 3 data sudah seimbang antara 0 dan 1. Hasil akhir dari proses SMOTE adalah sebuah data baru yang lebih seimbang tanpa mengurangi kualitas data asli.

5. Grid Search

Grid search adalah sebuah metode optimasi yang menjadikan titik-titik *grid* yang berjarak sama, lalu menghitung ukuran akurasi untuk setiap parameter tersebut sehingga ditemukan titik parameter yang paling optimal (titik yang memiliki ukuran akurasi terbaik) [11]. *Grid search* dalam penerapannya dikombinasikan dengan metode *cross validation*. *Cross validation* adalah sebuah metode pengembangan dari model *split validation* di mana validasinya mengukur *training error* dengan *test data* [12]. Dalam penentuan nilai *cross validation* direkomendasikan untuk pemilihan model terbaik adalah 10 [13].

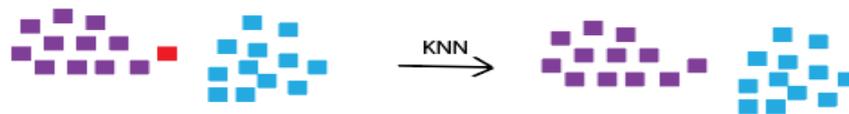


Gambar 4. Grid Search CV

Pada penelitian ini, nilai *cross validation* yang digunakan adalah 10 dengan range 1 - 30.

6. KNN

K-nearest neighbor adalah sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran (*neighbor*) yang jaraknya paling dekat dengan objek tersebut yang dihitung berdasarkan nilai *euclidean* [14]. Kasus khusus di mana klasifikasi diprediksi berdasarkan data pembelajaran yang paling dekat (*K*) disebut algoritma *nearest neighbor*, setelah nilai *K* ditentukan selanjutnya diambil 1 (*n*) tetangga lalu dilihat masing-masing dari tetangga tersebut apakah termasuk penderita kanker atau tidak [15].



Gambar 5. KNN

7. Evaluasi

Pada penelitian ini, untuk mengevaluasi model yang dibuat menggunakan *confusion matrix*. *Confusion matrix* adalah *matrix* yang menggambarkan performa dari sebuah model atau algoritma di mana setiap baris tersebut merepresentasikan kelas aktual dan setiap kolom merepresentasikan kelas prediksi [16].

| | | PREDIKSI | |
|--------|-----------------------|-----------------------|-----------------------|
| | | TN (True Negatif) | FP (False Positif) |
| ACTUAL | TN (True Negatif) | TN (True Negatif) | FP (False Positif) |
| | FN (False Negatif) | FN (False Negatif) | TP (True Positif) |

Gambar 6. *Confusion Matrix*

Dari Gambar 6 dapat digunakan untuk menentukan nilai *accuracy*, *recall*, *precision*, dan *F1-score*.

Accuracy adalah nilai seberapa sering model mengklasifikasikan dengan benar. Untuk mendapatkan nilai akurasi ditunjukkan pada Persamaan 1.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{Total}} \quad (1)$$

Precision adalah seberapa sering model memprediksi benar sesuai data prediksi. Untuk mendapatkan nilai *precision* ditunjukkan pada Persamaan 2.

$$\text{Precision} = \frac{\text{TP}}{\text{FP} + \text{TP}} \quad (2)$$

Recall adalah seberapa sering model memprediksi positif ketika kelas aktualnya positif. Untuk mendapatkan nilai *recall* ditunjukkan pada Persamaan 3.

$$\text{Recall} = \frac{\text{TP}}{\text{FN} + \text{TP}} \quad (3)$$

F1-score adalah rata-rata dari *precision* dan *recall*. Untuk mendapatkan nilai *F1-score* ditunjukkan pada Persamaan 4.

$$F1 - Score = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (4)$$

C. Hasil dan Pembahasan

1. Hasil Pelatihan dan Pengetesan Model

Sebelum memasuki tahap evaluasi menggunakan *confusion matrix*, harus dilakukan pemisahan antara data latih dan juga data uji. Pada penelitian ini, pembagian data latih sebesar 80% dari total data hasil SMOTE *oversampling*, sedangkan data uji sebesar 20% dari total data hasil SMOTE *oversampling*. Data latih juga dibagi menjadi dua, yaitu *x_train* sebagai kolom independen yang mempengaruhi target dan juga *y_train* sebagai kolom dependen yang digunakan sebagai target. Data latih digunakan untuk melatih algoritma KNN oleh sistem yang kemudian disimpan untuk digunakan dalam proses pelatihan.

Tabel 4. Data *x_train*

| No | <i>x_train</i> | Jumlah |
|--------------|-----------------------|-------------|
| 1 | YELLOW_FINGER | 432 |
| 2 | ANXIETY | 432 |
| 3 | PEER_PRESSURE | 432 |
| 4 | CHRONIC DISEASE | 432 |
| 5 | FATIGUE | 432 |
| 6 | ALLERGY | 432 |
| 7 | WHEEZING | 432 |
| 8 | ALCOHOL CONSUMING | 432 |
| 9 | COUGHING | 432 |
| 10 | SWALLOWING DIFFICULTY | 432 |
| 11 | CHEST PAIN | 432 |
| Total | | 4752 |

Tabel 5. Data *y_train*

| No | <i>y_train</i> | Jumlah |
|--------------|----------------|------------|
| 1 | LUNG_CANCER | 432 |
| Total | | 432 |

Tahap selanjutnya adalah melakukan uji data yang pada tahap berikutnya akan dievaluasi untuk didapatkan nilai akurasinya menggunakan metode *confusion matrix*.

Tabel 6. Data *x_test*

| No | <i>x_test</i> | Jumlah |
|----|-----------------|--------|
| 1 | YELLOW_FINGER | 108 |
| 2 | ANXIETY | 108 |
| 3 | PEER_PRESSURE | 108 |
| 4 | CHRONIC DISEASE | 108 |

| | | |
|--------------|-----------------------|-------------|
| 5 | FATIGUE | 108 |
| 6 | ALLERGY | 108 |
| 7 | WHEEZING | 108 |
| 8 | ALCOHOL CONSUMING | 108 |
| 9 | COUGHING | 108 |
| 10 | SWALLOWING DIFFICULTY | 108 |
| 11 | CHEST PAIN | 108 |
| Total | | 1188 |

Tabel 7. Data y_{test}

| No | y_{train} | Jumlah |
|--------------|-------------|------------|
| 1 | LUNG_CANCER | 108 |
| Total | | 108 |

2. Hasil Evaluasi Model

Tahap berikutnya setelah dilakukan pembagian data latih dan data uji, nilai K pada Algoritma KNN akan dioptimasi menggunakan *grid search cv* dengan *range* 1 – 30 untuk didapatkan nilai K yang optimal. Hasil dari optimasi ditampilkan pada Tabel 8.

Tabel 8. Hasil Optimasi

| No | Range | K | Accuracy | Precision | Recall | F1-Score |
|----|-------|----|----------|-----------|--------|----------|
| 1 | 26-30 | 27 | 0,93 | 0,93 | 0,93 | 0,93 |
| 2 | 21-25 | 21 | 0,94 | 0,94 | 0,94 | 0,93 |
| 3 | 16-20 | 17 | 0,93 | 0,93 | 0,93 | 0,93 |
| 4 | 11-15 | 11 | 0,94 | 0,95 | 0,94 | 0,94 |
| 5 | 6-10 | 9 | 0,95 | 0,96 | 0,95 | 0,95 |
| 6 | 1-5 | 3 | 0,96 | 0,97 | 0,96 | 0,96 |

Berdasarkan hasil Tabel 8, dapat dilihat perbandingan nilai *accuracy*, *precision*, *recall*, dan *F1-score* pada setiap nilai K yang diuji menggunakan metode *grid search cv*. Nilai K dengan *accuracy* tertinggi yaitu pada nilai 3 dengan nilai *accuracy* 0,96 atau 96%, yang artinya pada saat nilai K=3, KNN mampu mengklasifikasikan dengan tingkat kebenaran 96%. Nilai K=3 memiliki nilai *precision* sebesar 0,97 atau 97%, menunjukkan bahwa pada saat nilai K=3, KNN mampu memprediksi dengan tingkat kebenaran 97% sesuai hasil prediksi benar. Kemudian nilai K=3 juga memiliki nilai *recall* 0,96 atau 96%, yang menunjukkan bahwa pada saat nilai K=3, KNN mampu memprediksi positif pada saat kelas aktualnya positif dengan rasio 96%.

Nilai K dengan *accuracy* tertinggi kedua yaitu pada nilai 9 dengan nilai *accuracy* 0,95 atau 95%, yang artinya pada saat nilai K=9, KNN mampu mengklasifikasikan dengan tingkat kebenaran 95%. Nilai K=9 memiliki nilai *precision* sebesar 0,96 atau 96%, menunjukkan bahwa pada saat nilai K=9, KNN mampu memprediksi dengan tingkat kebenaran 96% sesuai hasil prediksi benar. Kemudian nilai K=9 juga memiliki nilai *recall* 0,95 atau 95%, yang menunjukkan bahwa pada saat nilai K=9, KNN mampu memprediksi positif pada saat kelas aktualnya positif dengan rasio 95%.

Nilai K dengan *accuracy* tertinggi ketiga yaitu pada nilai 11 dan 21 dengan nilai *accuracy* 0,94 atau 94%, yang artinya pada saat nilai K=11 atau 21, KNN mampu mengklasifikasikan dengan tingkat kebenaran 94%. Nilai K=11 memiliki nilai *precision* sebesar 0,95 atau 95%, sedangkan nilai K=21 memiliki nilai *precision* sebesar 0,94 atau 94%, menunjukkan bahwa pada saat nilai K=11, KNN mampu memprediksi dengan tingkat kebenaran 95% sesuai hasil prediksi benar dan pada saat nilai K=21 KNN mampu memprediksi dengan tingkat kebenaran 94% sesuai hasil prediksi benar. Kemudian nilai K=11 atau 21 memiliki nilai *recall* 0,94 atau 94%, yang menunjukkan bahwa pada saat nilai K=11 atau 21, KNN mampu memprediksi positif pada saat kelas aktualnya positif dengan rasio 94%.

Nilai K dengan *accuracy* tertinggi ketiga yaitu pada nilai 17 dan 27 dengan nilai *accuracy* 0,93 atau 93%, yang artinya pada saat nilai K=17 atau 27, KNN mampu mengklasifikasikan dengan tingkat kebenaran 93%. Nilai K=17 atau 27 memiliki nilai *precision* sebesar 0,93 atau 93%, menunjukkan bahwa pada saat nilai K=17 atau 27, KNN mampu memprediksi dengan tingkat kebenaran 93% sesuai hasil prediksi benar. Kemudian nilai K=17 atau 27 memiliki nilai *recall* 0,93 atau 93%, yang menunjukkan bahwa pada saat nilai K=17 atau 27, KNN mampu memprediksi positif pada saat kelas aktualnya positif dengan rasio 93%.

D. Simpulan

Berdasarkan penelitian dan pembahasan penerapan optimasi nilai k pada algoritma *k-nearest neighbor* menggunakan metode *grid search cv* dapat disimpulkan bahwa nilai k paling optimal yaitu 3 dengan perbandingan data latih 80% dan data uji 20%. Pada saat nilai k=3, algoritma KNN mampu memprediksi data dengan nilai akurasi 96%. Oleh karena itu, nilai k=3 sangat baik digunakan untuk memprediksi kanker menggunakan algoritma KNN.

E. Ucapan Terima Kasih

Terimakasih kepada Program Studi Informatika Fakultas Ilmu Komputer Universitas Amikom Yogyakarta yang telah memfasilitasi dan memberi bimbingan pada penelitian ini.

F. Referensi

- [1] World Health Organization, <https://www.iarc.who.int/news-events/lung-cancer-awareness-month-2022/> (accessed Aug. 1, 2023).
- [2] Global Cancer Observatory. (2021, March). *Indonesia Fact Sheets*. Internasional Agency for Research on Cancer.
- [3] Putra Kurniadi, R., Rohmat Saedudin, R., & Widartha, V. P. (n.d.). *PERBANDINGAN AKURASI ALGORITMA K-NEAREST NEIGHBOR DAN LOGISTIC REGRESSION UNTUK KLASIFIKASI PENYAKIT DIABETES*.
- [4] Madaerdo Sotarjua, L., Budhi Santoso, D., Singaperbangsa Karawang Jl Ronggo Waluyo, U. H., Telukjambe Timur, K., Karawang, K., & Barat, J. (n.d.). *PERBANDINGAN ALGORITMA KNN, DECISION TREE, DAN RANDOM FOREST PADA DATA IMBALANCED CLASS UNTUK KLASIFIKASI PROMOSI KARYAWAN*. 7(2), 2022.
- [5] Istadi, & Client. (2022). *Optimasi SVM menggunakan algoritme grid search untuk identifikasi citra biji kopi robusta berdasarkan circularity dan eccentricity*.

- <https://doi.org/10.14710/jtsiskom.2022.13807>
- [6] Veeralagan J. (2022). *Hyper Tuning Using Gridsearchcv on Machine Learning Models for Prognosticating Dementia*. <https://doi.org/10.21203/rs.3.rs-2316713/v1>
- [7] Thaufik Rizaldi, S. (2020). *Perbandingan Teknik Pembagian Data untuk Klasifikasi Sarana Akses Air pada Algoritma K-Nearest Neighbor dan Naïve Bayes Classifier* (Issue SNTIKI). <https://sarpras.dikdasmen.kemdikbud.go.id>
- [8] Al Aswad, N. (2022, April 22). Lung Cancer. <https://www.kaggle.com/datasets/nancyalasad90/lung-cancer>
- [9] Zulaikhah, S. H., Aziz, A., & Harianto, W. (2022). OPTIMASI ALGORITMA K NEAREST NEIGHBOR (KNN) DENGAN NORMALISASI DAN SELEKSI FITUR UNTUK KLASIFIKASI PENYAKIT LIVER. In *Jurnal Mahasiswa Teknik Informatika* (Vol. 6, Issue 2). <https://archive.ics.uci.edu/ml/index.php>
- [10] WIJAYANTI, N. P. Y. T., N. KENCANA, E., & SUMARJAYA, I. W. (2021). SMOTE: POTENSI DAN KEKURANGANNYA PADA SURVEI. *E-Jurnal Matematika*, 10(4), 235. <https://doi.org/10.24843/mtk.2021.v10.i04.p348>
- [11] Satriani, D., Khasanah, L. U., & Rizki, N. A. (2019). *Prosiding Seminar Nasional Matematika, Statistika, dan Aplikasinya*.
- [12] Muiz, A. (n.d.). Pengujian Data dengan Cross Validation. <https://www.pengalaman-edukasi.com/2020/04/apa-itu-k-fold-cross-validation.html>
- [13] Wibowo, antoni. (2017, November 24). *10 Fold-Cross Validation*. <https://mti.binus.ac.id/2017/11/24/10-fold-cross-validation/>
- [14] Ayu Fitria, st, Muslim, nd, & Huzain Azis, rd. (2018). Analisis Kinerja Sistem Klasifikasi Skripsi menggunakan Metode Naïve Bayes Classifier. *Prosiding Seminar Nasional Ilmu Komputer Dan Teknologi Informasi*, 3(2)
- [15] Angreni, I. A., Adisasmita, S. A., Ramli, M. I., & Hamid, S. (2019). PENGARUH NILAI K PADA METODE K-NEAREST NEIGHBOR (KNN) TERHADAP TINGKAT AKURASI IDENTIFIKASI KERUSAKAN JALAN. *Rekayasa Sipil*, 7(2), 63. <https://doi.org/10.22441/jrs.2018.v07.i2.01>
- [16] Widhi Saputro, I., & Wulan Sari, B. (2019). Uji Performa Algoritma Naïve Bayes untuk Prediksi Masa Studi Mahasiswa Naïve Bayes Algorithm Performance Test for Student Study Prediction. *Citec Journal*, 6(1).