

SENASTIKA Universitas Malikussaleh

DETEKSI DINI KANKER PARU PADA PASIEN MENGGUNAKAN METODE K-MEANS CLUSTERING

Dharma Bella Arianda*¹, Meisya Syahtira², Rahmatin Nisak*³, Thifal Salsabila⁴, Munirul Ula⁵

^{1,2,3,4,5} Teknik Informatika, Universitas Malikussaleh, Lhokseumawe, Aceh

Email: ¹dharna.210170265@mhs.unimal.ac.id, ²meisya.210170268@mhs.unimal.ac.id, ³rahmatin.210170269@mhs.unimal.ac.id, ⁴thifal.210170273@mhs.unimal.ac.id, ⁵munirulula@unimal.ac.id

Abstrak

Kanker paru-paru merupakan salah satu faktor utama penyebab kematian di seluruh dunia, dengan angka kematian yang tinggi karena keterlambatan dalam diagnosis. Gejala kanker paru sering kali baru terdeteksi pada tahap lanjut, sehingga deteksi dini sangat diperlukan untuk meningkatkan peluang kesembuhan. Penelitian ini bertujuan untuk mengembangkan metode deteksi dini kanker paru pada pasien menggunakan algoritma *K-Means Clustering* berdasarkan gejala yang dialami. Data yang digunakan mencakup 309 sampel pasien dengan 16 atribut yang mencerminkan kondisi kesehatan, seperti umur, kebiasaan merokok, batuk, dan sesak napas. Metode *K-Means* diterapkan untuk pengelompokan data pasien ke dalam beberapa klaster berdasarkan kesamaan gejala yang membantu dalam mengidentifikasi pola gejala yang terkait dengan risiko kanker paru-paru. Hasil penelitian menunjukkan bahwa algoritma ini berhasil membagi pasien ke dalam dua klaster utama yang mengindikasikan tingkat risiko yang berbeda. Penggunaan metode *K-Means Clustering* terbukti efektif dalam membantu diagnosis awal kanker paru-paru dan mampu memberikan dampak yang besar dalam meningkatkan akurasi deteksi serta mempercepat proses penanganan medis.

Kata Kunci: *Deteksi, Kanker Paru, K-Means, Cluster, Gejala*

Abstract

Lung cancer is one of the leading causes of death worldwide, with high mortality rates due to delays in diagnosis. Lung cancer symptoms are often only detected at an advanced stage, making early detection crucial to improving the chances of recovery. This study aims to develop an early detection method for lung cancer in patients using the K-Means Clustering algorithm based on the symptoms experienced. The data used consists of 309 patient samples with 16 attributes reflecting health conditions such as age, smoking habits, coughing, and shortness of breath. The K-Means method was applied to group patient data into several clusters based on symptom similarities, helping to identify symptom patterns associated with the risk of lung cancer. The results show that this algorithm successfully divided patients into two main clusters indicating different risk levels. The use of the K-Means Clustering method has proven effective in aiding early diagnosis of lung cancer and has the potential to significantly improve detection accuracy and accelerate the medical treatment process.

Keywords: *Detection, Lung Cancer, K-Means, Cluster, Symptom*

1. PENDAHULUAN

Paru-paru ialah komponen yang sangat penting dalam sistem pernapasan manusia dan memiliki peranan yang krusial dalam memenuhi kebutuhan oksigen tubuh. Selain itu, paru-paru juga berfungsi sebagai area utama untuk pertukaran oksigen dari udara yang kita hirup dengan karbon dioksida yang terdapat dalam darah [1]. Penyebab umum dari berbagai penyakit paru-paru biasanya terkait dengan menghirup udara yang terkontaminasi oleh debu, asap, serta virus dan bakteri yang dapat menyebabkan infeksi pada saluran pernapasan. Penyakit paru-paru ini dapat mempengaruhi individu dari berbagai rentang usia, mulai dari bayi yang baru lahir hingga orang dewasa yang lebih tua. Sangat penting untuk dicatat bahwa penyakit paru-paru seringkali sulit untuk disembuhkan dan memerlukan perhatian medis yang serius [3,14].

Kanker paru adalah salah satu jenis kanker yang memiliki tingkat kematian tertinggi di dunia. Tingginya angka kematian yang terkait dengan kanker paru sebagian besar disebabkan oleh keterlambatan diagnosis, di mana gejala baru muncul pada tahap lanjut ketika pengobatan menjadi lebih sulit. Oleh karena itu, deteksi dini kanker paru sangat penting untuk meningkatkan peluang kesembuhan dan menurunkan angka kematian. Metode ini dapat membantu dalam identifikasi pasien pada tahap awal, memungkinkan untuk melakukan tindakan medis yang lebih cepat dan tepat [2,4].

Penelitian oleh Hida Muhimmatul Husna (2023) dengan judul “Klasifikasi Penyakit Kanker Paru-Paru Berdasarkan Faktor Risiko Menggunakan Metode *K-Means*” menganalisis 1000 sampel pasien dengan 22 atribut.

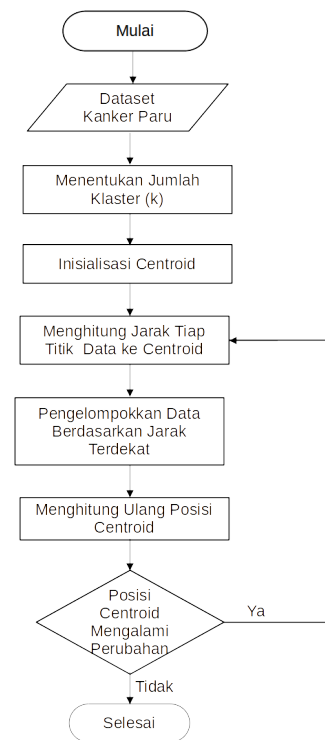
Metode *K-Means* digunakan setelah mengonversi data kategori menjadi numerik dan membagi data ke dalam rasio 90:10, 80:20, dan 70:30 untuk pelatihan dan pengujian. Evaluasi kinerja dilakukan menggunakan *confusion matrix* untuk menghitung *accuracy*, *precision*, dan *recall*. Hasil terbaik diperoleh pada rasio 90:10 dengan *accuracy* 73%, *precision* 73.61%, dan *recall* 73.51% [6].

Dalam penelitian ini, metode *K-Means Clustering* diterapkan dengan tujuan untuk mendeteksi secara dini kanker paru-paru pada pasien. Dengan memanfaatkan data yang telah tersedia, algoritma ini diharapkan dapat secara efektif mengelompokkan pasien berdasarkan berbagai gejala yang mereka miliki. Hal ini akan mempermudah proses dalam mendeteksi pasien yang berpotensi mengidap kanker paru-paru, sehingga memungkinkan intervensi yang lebih cepat dan tepat. Penelitian ini memiliki beberapa tujuan penting, yaitu untuk mengidentifikasi efektivitas metode *K-Means* dalam proses deteksi dini kanker paru-paru, serta mengembangkan model yang dapat digunakan untuk pendeteksian dini kanker paru-paru pada pasien dengan memanfaatkan metode *K-Means Clustering*. Model ini diharapkan dapat mengelompokkan pasien berdasarkan gejala yang ada, serta memberikan kontribusi yang signifikan dalam pengembangan alat bantu diagnosis berbasis data. Dengan demikian, diharapkan penelitian ini dapat membantu dalam upaya mengurangi angka kematian yang disebabkan oleh penyakit kanker paru-paru [5,11].

Penelitian ini penting karena kanker paru-paru merupakan salah satu faktor utama penyebab kematian di dunia, dan deteksi dini gejala sangat krusial untuk meningkatkan peluang kesembuhan. Dengan menggunakan metode *K-Means Clustering*, penelitian ini mampu mengelompokkan pasien berdasarkan gejala yang ada, sehingga memungkinkan identifikasi awal pada pasien yang berisiko. Pendekatan ini membantu dokter dalam memprioritaskan penanganan pasien yang menunjukkan gejala, meningkatkan akurasi diagnosis, serta memungkinkan intervensi medis yang lebih cepat, yang pada akhirnya dapat memperpanjang harapan hidup pasien dan mengurangi beban biaya perawatan [7,12].

2. METODE PENELITIAN

Penelitian ini mengadopsi metode dengan pendekatan *K-Means Clustering* untuk mendeteksi dini kanker paru-paru berdasarkan gejala pasien. Penelitian dimulai dengan pengumpulan data gejala dari pasien yang diperoleh dari *kaggle.com*. Dataset yang akan dianalisis sebanyak 309 sampel pasien dengan 16 atribut. Diantaranya *gender*, umur, merokok, jari menguning, kecemasan, penyakit kronis, kelelahan, alergi, batuk, dan lain-lain.



Gambar 1. Tahapan Penelitian

Langkah-langkah untuk menentukan *clustering* dengan metode *K-Means clustering*:

1. Menentukan jumlah kluster (k).

Langkah awal adalah menentukan jumlah kluster (k) yang diinginkan. Sebagai contoh, jika data akan dikelompokkan ke dalam tiga grup, maka nilai k ditetapkan sebagai 3.

2. Inisialisasi centroid
Pilih k titik secara acak dari data sebagai centroid awal. Titik-titik ini akan menjadi pusat awal dari masing-masing klaster.
3. Menghitung jarak tiap titik data ke centroid
Untuk setiap data, hitung jaraknya ke centroid dengan menggunakan metrik jarak seperti jarak *Euclidean*.
4. Pengelompokkan data berdasarkan jarak terdekat.
Setiap titik data dikelompokkan ke klaster yang memiliki centroid terdekat berdasarkan hasil perhitungan jarak.
5. Menghitung ulang posisi centroid.
Setelah semua data dikelompokkan, hitung ulang posisi centroid dengan menghitung rata-rata posisi dari semua titik data dalam klaster tersebut.
6. Mengulangi langkah 3-5
Langkah penghitungan jarak, pengelompokan data, dan penghitungan ulang centroid diulang hingga posisi centroid tidak lagi mengalami perubahan signifikan atau algoritma mencapai konvergensi [8].

Perhitungan jarak *Euclidean* menggunakan persamaan (1) :

$$d(x,y) = \sqrt{\sum_{k=1}^n (x_i - y_i)^2} \quad (1)$$

Keterangan:

d = Jarak antara x dan y

x = Data pusat *cluster*

y = Data pada atribut

n = Jumlah data

x_i = Data pada pusat *cluster* ke-i

y_i = Data pada setiap data ke-i

Perhitungan *cluster* dapat dilakukan dengan persamaan (2) :

$$C(i) = \frac{x_1 + x_2 + x_3 + \dots + x_n}{\sum x} \quad (2)$$

Keterangan:

x_1 = Nilai data *record* ke 1, dst

$\sum x$ = Jumlah data *record*

Pengelompokan data berdasarkan jarak terdekat menggunakan persamaan (3) :

$$Z = \frac{x - \tilde{x}}{\sigma} \quad (3)$$

Keterangan:

Z = Nilai standar

x = Data mentah

\tilde{x} = Nilai rata-rata

σ = Simpang baku

Dalam penelitian ini, *software* yang digunakan adalah *Visual Studio Code* dengan bahasa pemrograman *Python* menggunakan *Framework Flask*. *Visual Studio Code* adalah editor teks yang sangat populer di kalangan pengembang karena ringan dan mudah digunakan, mendukung berbagai bahasa pemrograman, termasuk *Python*. *Python* sendiri adalah bahasa pemrograman yang terkenal karena kesederhanaannya dan banyak digunakan untuk berbagai keperluan, seperti pengembangan *web*, analisis data, hingga kecerdasan buatan. Salah satu *framework Python* yang sering digunakan untuk membuat aplikasi *web* adalah *Flask*. *Flask* dikenal karena ringan, mudah dipelajari dan fleksibel, sehingga cocok untuk membuat aplikasi *web* mulai dari yang sederhana hingga yang lebih kompleks [9,15].

3. HASIL DAN PEMBAHASAN

Pada bagian ini, akan disajikan hasil dari penerapan metode K-Means Clustering yang dirancang khusus untuk mendeteksi secara dini kanker paru-paru pada pasien. Dalam proses ini, perhatian akan diberikan pada berbagai gejala yang dimiliki oleh masing-masing individu, sehingga memungkinkan analisis yang lebih mendalam dan akurat. Analisis hasil penelitian ini akan difokuskan pada pengelompokan pasien sesuai dengan-

kategori yang dihasilkan oleh algoritma K-Means, serta melakukan interpretasi mendalam terhadap setiap klaster yang terbentuk. Selain itu, pembahasan juga akan mencakup evaluasi kinerja model berdasarkan akurasi pengelompokan yang dihasilkan dan relevansinya terhadap kondisi medis yang dialami oleh pasien.

Setelah melalui proses pengolahan data dan penerapan algoritma *K-Means Clustering*, pasien kemudian dibagi ke dalam beberapa klaster yang didasarkan pada kemiripan gejala yang muncul. Setiap klaster tersebut merepresentasikan kelompok pasien yang memiliki karakteristik serupa dalam hal gejala yang dialami, dengan tingkat risiko yang berbeda-beda untuk mengidap kanker paru-paru. Dalam hal ini, terdapat klaster tertentu yang menunjukkan potensi yang lebih tinggi untuk mengidap kanker paru-paru pada tahap awal, sehingga memberikan informasi berharga yang dapat digunakan untuk penanganan lebih lanjut.

Dari hasil pengelompokan, terlihat bahwa metode *K-Means Clustering* mampu mengidentifikasi pola gejala yang relevan dengan risiko kanker paru. Klaster-klaster yang terbentuk menunjukkan bahwa gejala seperti merokok, batuk, dan sesak napas lebih dominan pada pasien yang masuk dalam klaster berisiko tinggi. Hal ini menunjukkan bahwa metode ini dapat diandalkan untuk mendukung deteksi dini pasien dengan potensi kanker paru-paru.

3.1 Tampilan Menu Utama

Tampilan Menu Utama dapat dilihat pada Gambar.



Gambar 2. Tampilan Menu Utama

Gambar di atas menunjukkan tampilan awal dari antarmuka *web* sebelum *file CSV* diunggah untuk dianalisis menggunakan metode *K-Means Clustering*. Pada halaman ini, pengguna diminta untuk mengunggah data *CSV* melalui tombol *Choose File*, diikuti dengan menekan tombol *Submit* untuk memulai proses pengelompokan. *Web* ini digunakan untuk mendeteksi dini kanker paru dengan melakukan *clustering* pada data gejala pasien. *Web* tersebut dirancang menggunakan *software Visual Studio Code* dengan bahasa pemrograman *Python* dan *framework Flask*.

3.2 Hasil Data Awal dan Hasil Data Akhir

Hasil K-Means Clustering

No	GENDER	AGE	SMOKING	YELLOW_FINGERS	ANXIETY	PEER_PRESSURE	CHRONIC DISEASE		FATIGUE	ALLERGY	WHEEZING	ALCOHOL CONSUMING	COUGHING	SHORTNESS OF BREATH	SWALLOWING DIFFICULTY	CHEST PAIN	Cluster
1	M	69	1	2	2	1	1	2	1	2	2	2	2	2	2	2	2
2	M	74	2	1	1	1	2	2	2	1	1	1	2	2	2	2	2
3	F	59	1	1	1	2	1	2	1	2	1	2	2	1	2	1	1
4	M	63	2	2	2	1	1	1	1	1	2	1	1	2	2	1	1
5	F	63	1	2	1	1	1	1	1	2	1	2	2	1	1	1	1
6	F	75	1	2	1	1	2	2	2	2	1	2	2	1	1	1	2
7	M	52	2	1	1	1	1	2	1	2	2	2	2	1	2	1	1
8	F	51	2	2	2	2	1	2	2	1	1	1	2	2	1	1	1
9	F	68	2	1	2	1	1	2	1	1	1	1	1	1	1	1	2
10	M	53	2	2	2	2	2	1	2	1	2	1	1	2	2	1	1
11	F	61	2	2	2	2	2	2	1	2	1	2	2	2	1	1	1
12	M	72	1	1	1	1	2	2	2	2	2	2	2	1	2	2	2
13	F	60	2	1	1	1	1	2	1	1	1	1	2	1	1	1	1
14	M	58	2	1	1	1	1	2	2	2	2	2	2	1	2	1	1

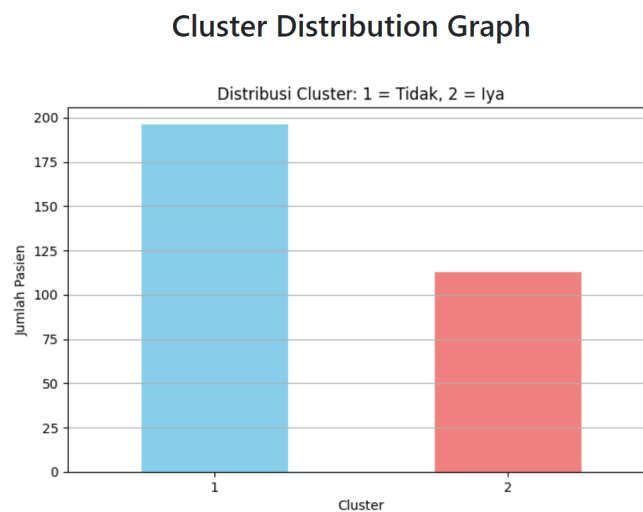
Gambar 3. Hasil Data Awal

298	F	47	2	2	1	2	2	2	2	2	1	2	2	1	1	1
299	M	62	2	1	2	1	1	2	1	2	2	2	2	1	2	1
300	M	65	2	2	2	2	1	2	2	1	1	1	2	2	1	2
301	F	63	2	2	2	2	2	2	2	2	1	2	2	2	2	1
302	M	64	1	2	2	2	1	1	2	1	2	1	1	2	2	1
303	F	65	2	2	2	2	1	2	1	2	1	2	2	2	1	2
304	M	51	1	2	1	1	2	2	2	2	2	2	2	1	2	1
305	F	56	1	1	1	2	2	2	1	1	2	2	2	2	1	1
306	M	70	2	1	1	1	1	2	2	2	2	2	2	1	2	2
307	M	58	2	1	1	1	1	2	2	2	2	2	1	1	2	1
308	M	67	2	1	2	1	1	2	2	1	2	2	2	1	2	2
309	M	62	1	1	1	2	1	2	2	2	2	1	1	2	1	1

Gambar 4. Hasil Data Akhir (Lanjutan)

Gambar di atas menunjukkan hasil *clustering* menggunakan metode *K-Means*, yang diterapkan pada data pasien dengan berbagai gejala kesehatan. Setiap pasien diidentifikasi dengan nilai numerik pada setiap variabel, yang mewakili kondisi atau gejala yang dialami. Pada gambar di atas terlihat hasil klasterisasi yang telah dilakukan. Terdapat dua *cluster* yaitu *cluster 1* (negatif) dan *cluster 2* (positif).

3.3 Grafik Clustering



Gambar 5. Hasil Grafik Clustering

Gambar di atas menunjukkan grafik *clustering* yang dihasilkan dari penerapan metode *K-Means Clustering*, menggunakan *software Visual Studio Code* dengan bahasa pemrograman *Python* dan *framework Flask*. Grafik ini menggambarkan jumlah pasien yang terkelompok ke dalam dua *cluster*, yaitu

- *Cluster 1* mewakili kelompok "Tidak" atau negatif, sedangkan *Cluster 2* mewakili kelompok "Iya" atau positif, dengan masing-masing kategori mengacu pada kondisi atau gejala yang teridentifikasi pada pasien.
- Sumbu x menunjukkan klaster yang terbentuk, sementara sumbu y menampilkan jumlah pasien.
- Grafik ini mengindikasikan bahwa jumlah pasien dalam *Cluster 1* lebih besar dibandingkan *Cluster 2*, yang menunjukkan data pasien lebih banyak memiliki gejala atau kondisi yang tergolong "Tidak" atau Negatif dalam konteks kanker paru.
- Hasil grafik menunjukkan bahwa dari 309 sampel pasien, terdapat 113 pasien yang terdiagnosis menderita kanker paru-paru dan 196 pasien lainnya tidak mengidap penyakit tersebut.

4. DISKUSI

Hasil penelitian ini menunjukkan bahwa metode *K-Means Clustering* berhasil mengelompokkan pasien kanker paru menjadi dua klaster yaitu, klaster pertama mengelompokkan data pasien dengan hasil negatif (tidak), sedangkan klaster kedua mengelompokkan data pasien dengan hasil positif (iya) terkait deteksi kanker paru berdasarkan gejala yang muncul. Gejala seperti merokok, batuk, dan sesak napas lebih dominan pada pasien di klaster risiko tinggi, mendukung bahwa metode ini dapat membantu mendeteksi dini kanker paru [13].

Penelitian ini sejalan dengan studi Hida Muhimmatul Husna (2023) yang juga menggunakan *K-Means* untuk mengelompokkan pasien berdasarkan gejala. Namun, penelitian ini lebih berfokus pada deteksi dini kanker paru, sedangkan penelitian terdahulu mengkaji penyakit kronis secara umum.

Salah satu kelemahan utama *K-Means* adalah sulitnya menentukan jumlah klaster yang optimal [10]. Jika jumlah klaster tidak tepat, pengelompokan bisa kurang akurat. Selain itu, metode ini sangat bergantung pada inisialisasi awal centroid. Hasil clustering dapat bervariasi jika centroid tidak dipilih dengan benar, seperti yang juga ditemukan dalam penelitian sebelumnya. Tantangan lain adalah keterbatasan literatur terkait penggunaan *K-Means* khusus untuk kanker paru.

5. KESIMPULAN

Kesimpulan dari penelitian ini menunjukkan bahwa metode *K-Means Clustering* mampu mendeteksi dini kanker paru berdasarkan gejala yang dialami pasien. Dengan mengelompokkan pasien ke dalam klaster berdasarkan kemiripan gejala, metode ini berhasil mengidentifikasi pola gejala yang relevan dengan risiko kanker paru. Hasil analisis menggunakan metode *K-Means Clustering* menunjukkan bahwa dari 309 sampel pasien, terdapat 113 pasien yang terdiagnosis menderita kanker paru-paru dan 196 pasien lainnya tidak mengidap penyakit tersebut. Klaster yang terbentuk memperlihatkan bahwa gejala seperti merokok, batuk, dan sesak napas lebih dominan pada pasien berisiko tinggi. Hasil ini menegaskan bahwa *K-Means Clustering* dapat menjadi alat yang andal dalam mendukung deteksi dini kanker paru.

DAFTAR PUSTAKA

- [1] Alfarisa, Sellyn, Efriza, Sri Wahyuni. "Karakteristik Pasien Kanker Paru di RSUP Dr. M. Djamil Padang Tahun 2021". *SCIENA*. Vol II No 6. 2023.
- [2] Alimin, Mohd. Ni Putu Rita Jeniyanti. "Pengaruh Penggunaan Fiksasi Masker Paru Terhadap Ketepatan Target Penyinaran Pada Kanker Paru Teknik Intensity Modulated Radiation Therapy (IMRT) Di Departemen Onkologi Radiasi Rumah Sakit Umum Jakarta". Vol.2, No.1. Hal 216-224. Januari 2024.
- [3] Binarsa, Bagus Deka. Citra Paramita Esti Cahyaningrum. "Perbedaan Patologi Anatomi Tingkat Kerusakan Alveoli Paru dengan Paparan Asap Rokok Konvensional dan Rokok Elektrik". Vol. 4. No. 1. hlm. 29-36. November 2022.
- [4] Buana, Indra. Dwi Agustian Harahap. "Asbestos, Radon Dan Polusi Udara Sebagai Faktor Resiko Kanker Paru Pada Perempuan Bukan Perokok". Vol.8 No.1 Mei 2022.
- [5] Fadilah Nur Siti, dkk. "Pengaruh Reduksi Fitur Pada Klasifikasi Kanker Paru Menggunakan CNN Dengan Arsitektur GoogLeNet". Vol. 12, No. 1. April 2023.
- [6] Husna. Hida Muhimmatul. "Klasifikasi Penyakit Kanker Paru-Paru Berdasarkan Faktor Risiko Menggunakan Metode K-Means". Universitas Islam Negeri Maulana Malik Ibrahim. 2023.
- [7] Julita Sinurat. "Jaringan Saraf Tiruan Diagnosa Penyakit Kanker Paru-Paru Menggunakan Metode Hebb Rule". Vol. 2, No 1. Page 20 - 27. Maret 2021.
- [8] Kurniawan Dwi Ferry Sri Dianova, Rinaldy Rinaldy. "Analisis Keterlambatan Diagnosis dan Terapi Kanker Paru di RSUD Dr. Zainoel Abidin, Banda Aceh". Vol. 5, No. 1, Hlm. 1- 7, April 2024.
- [9] M.T.I.Sambi Ua, Angelina. dkk. "Penggunaan Bahasa Pemrograman Python Dalam Analisis Faktor Penyebab Kanker Paru-Paru". *Jurnal Publikasi Teknik Informatika (JUPTI)*. Vol.2, No.2. Hal 88-99. Mei 2023.
- [10] Maiyanti Indra Sri, dkk. "Perbandingan Klasifikasi Penyakit Kanker Paru-paru menggunakan Support Vector Machine dan K-Nearest Neighbor". VOL. 18, No. 1. April 2023.
- [11] Muhlisoh, Asni Hasaini. "Hubungan Pengetahuan Laki-Laki Usia Middle Age Tentang Kanker Paru Dengan Kebiasaan Merokok". Vol. 5 No. 2 Oktober, 2021.
- [12] Rofiani, Riska. dkk. "Penerapan Metode Klasifikasi Decision Tree dalam Prediksi Kanker Paru-Paru Menggunakan Algoritma C4.5". *Jurnal TEKNO KOMPAK*, Vol. 18, No. 1. Hal. 126-139.
- [13] Salsabila Syifa, Shinta Ayu Intan, Dewi Wahyu Fitriana. "Gambaran Tipe Sel Karsinoma Paru Primer Berdasarkan Usia, Jenis Kelamin, dan Paparan Rokok di RSUP Dr. M. Djamil Padang Tahun 2018-2020". VOL - 04 No. 04 2023.

- [14] Sari Laura, Annisa Romadloni, Rostika Listyaningrum, "Penerapan Data Mining dalam Analisis Prediksi Kanker Paru Menggunakan Algoritma Random Forest," Vol.14, No.01, pp.155-162. Januari 2023.
- [15] Septhya Dhini. dkk. "Implementasi Algoritma Decision Tree dan Support Vector Machine untuk Klasifikasi Penyakit Kanker Paru". Vol. 3. pp: 15-19. 1 April 2023.