

# Penerapan Algoritma *Machine Learning* Dalam Mengklasifikasi Data Masa Studi di Indonesia Berdasarkan Jenis Kelamin

Aryanata Wijaya<sup>1</sup>, Waeisul Bismi<sup>2</sup>

<sup>1,2</sup>Jurusan Informatika Fakultas Teknik dan Informatika Universitas Bina Sarana Informatika

[115200054@bsi.ac.id](mailto:115200054@bsi.ac.id)

[waeisul.wbn@bsi.ac.id](mailto:waeisul.wbn@bsi.ac.id)

**Abstrak**— Masalah perbedaan jenis kelamin dan pendidikan di Indonesia adalah masalah yang serius, karena dengan pendidikan yang layak barulah Negara Indonesia ini bisa menjadi Negara yang maju. Untuk menunjang penelitian ini maka penulis mengambil data rata-rata lama sekolah dari Badan Pusat Statistik sebanyak 1.028 data kemudian diberikan pelabelan Lama dan Sebentar serta dibagi menjadi 4 kategori yaitu Data Saat Pandemi *Covid* Laki-laki dan Perempuan, dan Data Pasca *Covid* Laki-laki dan Perempuan selanjutnya diolah dengan bantuan aplikasi *Orange Data mining*, selanjutnya data dirapuhkan menggunakan *Data Table* dan dikelompokkan atributnya menggunakan *Select Column* dan diproses menggunakan algoritma *K-Nearest Neighbor*, serta menggunakan *Test and Score* dan *Confusion Matrix* untuk melihat hasil evaluasinya. Setelah data tersebut diolah didapatkan hasil dari pengolahan data tersebut yaitu laki-laki lebih lama masa studi atau sekolahnya dibandingkan perempuan dari data Saat Pandemi *Covid* dan Pasca Pandemi *Covid* serta memperoleh tingkat akurasi dari 4 kategori yaitu 0.909, 0.848, 0.846, 0.920 dengan menggunakan algoritma *K-Nearest Neighbor*, sehingga pengolahan data menghasilkan hasil yang optimal. Dengan dilakukannya penelitian ini agar dapat menjadi referensi bagi pemerintah Indonesia agar dapat memajukan pendidikan di Indonesia, karena Negara yang maju dan berkualitas diwujudkan dari Sumber Daya Manusia yang berkualitas juga dan itu didapatkan dari Pendidikan yang bagus, serta dapat menjadi referensi bagi penelitian selanjutnya.

**Kata Kunci**— Jenis Kelamin, *K-Nearest Neighbor*, Pendidikan di Indonesia, Pandemi *Covid*, Rata-rata Lama Sekolah.

## I. PENDAHULUAN

Masalah kesetaraan *gender* atau jenis kelamin di Indonesia bukan lah suatu masalah yang tabu, karena masalah ini sudah sangat sering terjadi, seperti perempuan tidak perlu sekolah tinggi-tinggi karena akan menjadi ibu rumah tangga, wanita tidak boleh melakukan hal-hal yang biasa laki-laki lakukan dan lain sebagainya sering kali wanita yang menjadi sasarannya maka dari itu untuk dapat mengatasi masalah ini diperlukan pendidikan yang berkualitas, kemudian masalah pendidikan di Indonesia juga patut kita perhatikan dimana banyak sekali masyarakat di Indonesia yang putus sekolah dan tidak sekolah dengan berbagai macam alasan seperti dilarang orang tua, tidak ada biaya, dari rumah ke sekolah jauh, dan lain sebagainya sehingga dapat berdampak negatif dari individu itu sendiri dan lingkungannya, dari masalah-masalah diatas membuat penulis prihatin dan ingin berkontribusi didalamnya lewat penelitian penulis ini. Didalam penelitian [1] menjelaskan bahwa, Tinggi rendahnya kualitas suatu negara ditentukan juga oleh peran

rakyatnya dalam pembangunan negara tersebut, dan pendidikan merupakan salah satu prasyarat pembangunan suatu negara untuk menghasilkan sumber daya manusia yang berkualitas.

Sedangkan menurut Siswati, Endang & dkk didalam penelitian [2] menjelaskan bahwa :

“Dimensi pemerataan dalam pendidikan merupakan faktor penting ketika mempertimbangkan upaya dalam mensejahterakan masyarakat dalam pembangunan negri ini. Kita berhipotesis bahwa semakin berkualitas pendidikan masyarakat suatu daerah, maka semakin berkualitas pula manusianya baik dalam hal pola pikir maupun perilakunya (Siswati, Endang & dkk., 2018). Kesetaraan jenis kelamin dalam dunia pendidikan harus ditingkatkan lebih lanjut, khususnya bagi perempuan”.

Sumber Daya Manusia (SDM) yang baik terlahir dari pondasi yang kuat, dan pondasinya adalah dari Sekolah atau Pendidikan, dari Pendidikan yang bermutu menghasilkan sumber daya manusia yang berkualitas serta itu nantinya dapat bermanfaat bagi Negara, SDM yang berkualitas tentunya akan memberikan dampak yang positif, seperti berkurangnya angka pengangguran, kriminalitas dan juga kepribadian setiap individu dapat lebih baik lagi, pentingnya pendidikan harus sering dikampanyekan seperti orang tua yang berpikir bahwa sekolah hanya buang-buang uang, lebih baik langsung kerja dan dapat uang, hal ini juga yang harus jadi perhatian lebih oleh pemerintah.

Kita bisa melihat berapa banyak wanita atau laki-laki yang sekolah di masing-masing kota, dan bagaimana performa pendidikan di masa pandemi dan setelah pandemi, data-data yang ada pastinya banyak dan untuk mengolahnya dibutuhkan bantuan *Machine Learning*, data yang banyak ini dapat diolah menjadi data yang terstruktur, mudah untuk ditinjau dan diklasifikasikan, juga dapat nantinya digunakan oleh Pemerintah untuk meningkatkan dan meratakan pendidikan di Indonesia sampai ke pelosok-pelosok daerah karena penulis berharap Indonesia dapat semakin maju dalam segala hal dan pendidikan adalah hal yang utama dan harus sangat-sangat diperhatikan.

## II. LANDASAN TEORI

### A. Tinjauan Pustaka

#### 1) Machine Learning

Didalam penelitian [3] menjelaskan bahwa, Pembelajaran mesin adalah bagian dari bidang kecerdasan buatan yang memiliki fokus untuk mengembangkan algoritma dan model komputer yang bisa melakukan pembelajaran dari suatu model data yang dapat melakukan tugas tertentu tanpa memerlukan pemrograman yang sulit. Sedangkan didalam penelitian [4] menjelaskan bahwa, *Machine Learning* digunakan untuk mengkerucutkan cabang ilmu komputer yang mempelajari tentang cara mendesain algoritma yang dapat mempelajari atau menyesuaikan dengan pola data dan tidak perlu diprogram lagi. Dan didalam penelitian [5] menjelaskan bahwa, *Machine Learning* adalah suatu aplikasi komputer dan algoritma matematika yang cara pembelajaran berasal dari data dan memunculkan prediksi di masa depan.

Sehingga dapat saya simpulkan bahwa, *Machine Learning* adalah sebuah program aplikasi berbasis Kecerdasan Buatan yang dapat belajar sendiri dari data yang disediakan dan dapat menghasilkan suatu keputusan yang tidak perlu di kendalikan oleh manusia.

#### 2) K-Nearest Neighbor (K-NN)

Didalam penelitian [6] menjelaskan bahwa, Algoritma *K-Nearest Neighbor* adalah sebuah metode yang dipakai untuk mengklasifikasi atribut dan data sampel pelatihan dan sering kali memberikan hasil yang signifikan dibandingkan metode *data mining* lainnya. Sedangkan didalam penelitian [7] menjelaskan bahwa, K-NN menangani masalah klasifikasi dengan menentukan jarak antara sampel baru yang akan diklasifikasikan dengan sampel pelatihan, kemudian mengamati sampel baru dan memutuskan apakah sampel baru termasuk ke dalam kelas yang paling banyak sampel baru.

Sehingga dapat saya simpulkan bahwa, *K-Nearest Neighbors* (K-NN) adalah algoritma yang dapat mengatasi masalah klasifikasi dengan cara mengamati data sampel baru dan menerapkannya untuk mengklasifikasi.

#### 3) Klasifikasi

Didalam penelitian [8] menjelaskan bahwa, Klasifikasi merupakan proses menilai suatu data berdasarkan inputan data ke kategori tertentu. Dan tujuannya adalah untuk memperoleh prediksi dari suatu kategori berdasarkan objek kategori tersebut yang tidak diketahui karakternya. Sedangkan didalam penelitian [9] menjelaskan bahwa, Klasifikasi merupakan pembelajaran terhadap algoritma klasifikasi yang melakukan penginputan data latih untuk dasar *system* melakukan proses klasifikasi untuk data uji, dan dapat melakukan pengkategorian ciri tertentu.

Sehingga dapat saya simpulkan bahwa, Klasifikasi adalah suatu proses pembelajaran dengan memasukkan data kedalam kelas dan melakukan pengujian terhadap data tersebut sehingga menghasilkan pengelompokan data.

#### 4) Data Collection

Didalam penelitian [10] menjelaskan bahwa, *Data Collection* merupakan tahap pengumpulan serta memastikan *variable of interest* secara sistematis, dan dalam melakukan *Data Collection* memiliki standarisasi berupa ekstraksi data dan pemeriksaan jurnal yang dilakukan oleh semua anggota penelitian. Sedangkan didalam penelitian [11] bahwa, *Data Collection* adalah proses mengumpulkan, dan menganalisis informasi menggunakan cara tertentu. Dan tujuannya adalah untuk mengumpulkan data agar dapat membantu dalam membuat keputusan. Dan didalam penelitian [12] menjelaskan bahwa, *Data Collection* adalah cara yang digunakan peneliti untuk memperoleh data, dan data tersebut bisa didapatkan melalui, kuesioner, studi kasus, observasi, investigasi, dokumentasi dan lain-lain.

Sehingga dapat saya simpulkan bahwa, *Data Collection* adalah proses pengumpulan data agar dapat dianalisis sesuai dengan kebutuhan penelitian.

#### 5) Labeling

Didalam penelitian [13] menjelaskan bahwa, *Labeling* adalah tahapan untuk memberikan tanda atau klasifikasi untuk data-data berdasarkan kategori atau klasifikasi yang telah diputuskan sebelumnya. Dan pada konteks *data mining*, *labeling* juga dilakukan untuk menentukan (positif atau negatif). Sedangkan didalam penelitian [14] menjelaskan bahwa, *Labeling* biasanya melibatkan manusia, sehingga banyak alat pelabelan telah dibangun untuk membantu *melabeli* objek dari berbagai jenis untuk berbagai tugas pembelajaran (misalnya, klasifikasi, segmentasi, dan lain-lain).

Sehingga dapat saya simpulkan bahwa, *Labeling* adalah proses memberikan *label* atau tanda untuk suatu objek atau data yang berfungsi untuk memudahkan proses klasifikasi, segmentasi, dan lainnya.

#### 6) Data Table

Didalam penelitian [15] menjelaskan bahwa, *Data Table* adalah fitur yang dapat digunakan untuk melihat data yang akan diproses, kemudian menerima satu atau beberapa set data sebagai masukan dan menjabarkannya menjadi *spreadsheet*. Contohnya data bisa diurutkan dari nilai atributnya. *Data Table* ini juga memungkinkan memilih contoh data secara manual.

#### 7) Select Column

Didalam penelitian [16] menjelaskan bahwa, *Select Columns* digunakan untuk memilih kolom yang akan digunakan. Pengguna dapat memutuskan atribut mana yang akan digunakan dan bagaimana caranya dan membedakan antara atribut biasa, atribut kelas (opsional), dan atribut *Meta*.

#### 8) Cross Validation

Menurut Raschka didalam penelitian [17] menjelaskan bahwa, *Cross Validation* adalah proses membagi data hingga dua bagian yaitu *data training* dan *data testing* secara terus menerus, dan setiap data mempunyai kesempatan yang sama untuk kembali dibagi menjadi 2 kategori data. Raschka (2018). Sedangkan didalam penelitian [18] menjelaskan bahwa, *Cross Validation* adalah cara yang digunakan untuk teknik *data mining* yang memperoleh hasil akurasi penuh ketika data dibagi menjadi dua *subset* (*data training* dan *data testing*). Dan didalam penelitian [19] menjelaskan bahwa, *Cross Validation*

adalah cara untuk menilai sejauh mana hasil analisa statistik untuk *Dataset* yang secara mandiri terpisah, utamanya dapat digunakan untuk memprediksi dan mengevaluasi tingkat akurasi dari model prediksi dan diterapkan dalam situasi dunia nyata.

Sehingga dapat saya simpulkan bahwa, *Cross Validation* adalah suatu metode untuk membagi 2 data yaitu *training* dan *testing* secara berulang sampai menghasilkan hasil yang akurat dan dapat diterapkan.

#### 9) Test and Score

Didalam penelitian [15] menjelaskan bahwa, *Test and Score* adalah metode untuk menguji algoritma pembelajaran dan tersedia berbagai skema pengambilan sampel, menampilkan tabel dengan ukuran kinerja pengklasifikasi yang berbeda, serta mengeluarkan hasil yang dapat dipakai untuk menganalisa performa pengklasifikasian lainnya.

#### 10) Confusion Matrix

Didalam penelitian [20] menjelaskan bahwa, *Confusion Matrix* adalah untuk mengukur performa untuk masalah klasifikasi *Machine Learning* yang hasilnya didapat berupa dua kelas atau lebih. serta membagikan perbandingan hasil klasifikasi yang dilakukan sistem dan memperoleh hasil klasifikasi yang sebenarnya. Sedangkan didalam penelitian [21] menjelaskan bahwa, *Confusion Matrix* adalah suatu pengukuran berbentuk matrik 2x2 yang dapat digunakan untuk memperoleh banyak ketepatan algoritma yang dipakai. Dan didalam penelitian [22] menjelaskan bahwa, *Confusion Matrix* adalah *table* dengan empat kombinasi yang berbeda dari prediksi dan aktual. Ada empat istilah yang merupakan hasil dari proses klasifikasi pada *Confusion Matrix* yaitu *True Positive*, *True Negative*, *False Positive*, dan *False Negative*. Seperti dalam *table* I dibawah ini :

Tabel I  
*Confusion Matrix*

		<i>Predicted Values</i>	
		<i>Positive</i>	<i>Negative</i>
<i>Actual Values</i>	<i>Positive</i>	<i>TP (True Positive)</i>	<i>FP (False Positive)</i>
	<i>Negative</i>	<i>FN (False Negative)</i>	<i>TN (True Negative)</i>

Sumber : [22]

Penjelasan dari *table* diatas adalah :

1. *TP (True Positive)* : Suatu data yang diprediksikan sebagai *positive* dan hasilnya benar *positive*.
2. *FN (False Negative)* : Suatu data yang diprediksikan sebagai *negative* dan hasilnya ternyata *positive*.
3. *FP (False Positive)* : Suatu data yang diprediksikan sebagai *positive* dan hasilnya ternyata *negative*.
4. *TN (True Negative)* : Suatu data yang diprediksikan sebagai *negative* dan hasilnya benar *negative*.

Sehingga dapat saya simpulkan bahwa, *Confusion Matrix* adalah alat yang berfungsi untuk mengukur mengukur performa dari hasil klasifikasi *Machine Learning*.

#### 11) Orange Data mining

Didalam penelitian [23] menjelaskan bahwa, *Orange Data mining* adalah aplikasi yang memungkinkan pembuatan alur

kerja pemrosesan data standar untuk berbagai masalah pembelajaran mesin. *Orange Data mining* juga berada dalam *Open Source Domain*. Sedangkan didalam penelitian [24] menjelaskan bahwa, *Orange Data mining* adalah seperangkat alat pembelajaran mesin, penambahan data, dan skrip *Python* yang dikembangkan untuk analisis data interaktif dan konstruksi metode penambahan data berbasis komponen.

Sehingga dapat saya simpulkan bahwa, *Orange Data mining* adalah sebuah *software* yang digunakan untuk membantu pengolahan data, teks dan lainnya.

#### B. Penelitian Terkait

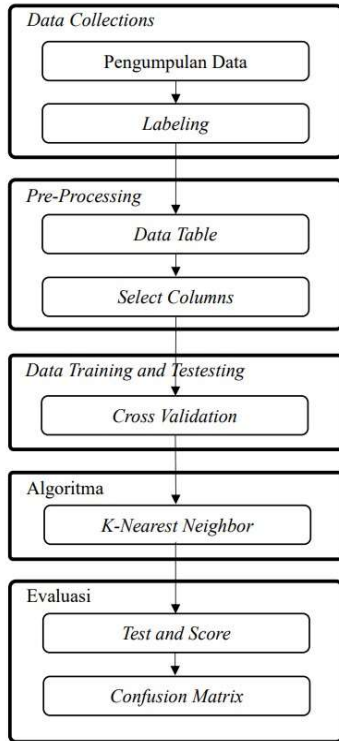
Penelitian ini bertujuan untuk Mengklasifikasi data masa studi di Indonesia berdasarkan jenis kelamin dengan memakai algoritma *K-Nearest Neighbor* (K-NN), dan dalam penelitian terkait ini penulis merujuk pada penelitian terkait atau penelitian terdahulu sebagai bahan acuan yang dapat dilihat pada rangkuman dari penelitian terkait sebagai berikut :

1. Penelitian yang pertama dilakukan oleh [25] dengan judul “Implementasi Klasifikasi *Naive Bayes* Dalam Memprediksi Lama Studi Mahasiswa (Studi Kasus : Universitas Dhyana Pura)”. Penelitian dilakukan untuk memprediksi berapa lama masa studi mahasiswa dengan menggunakan metode *Naive Bayes*, dan hasil yang didapatkan adalah penelitian ini dapat mencapai akurasi 100% dengan mengimplementasikan *data mining* menggunakan *Naive Bayes*, yang dibagi menjadi 968 data pelatihan dan 193 data pengujian.
2. Penelitian yang kedua dilakukan oleh [26] dengan judul “Analisis Penggunaan Algoritma Klasifikasi Dalam Prediksi Kelulusan Menggunakan *Orange Data mining*”. Penelitian dilakukan untuk mengklasifikasi dan memprediksi kelulusan dengan menggunakan 3 metode yaitu *Naive Bayes*, K-NN dan *Neural Network*, dan hasil yang didapatkan adalah membandingkan ketiga metode tersebut, dan mendapatkan metode terbaik yaitu metode K-NN dengan nilai akurasi 89%.
3. Penelitian ini dilakukan oleh [27] dengan judul “Penerapan Model Klasifikasi *Naive Bayes* Untuk Prediksi Mahasiswa Tepat Waktu”. Penelitian dilakukan untuk memprediksi sejumlah mahasiswa yang tepat waktu dengan menggunakan metode *Naive Bayes*, dan hasilnya algoritma *Naive Bayes* menghasilkan akurasi sebesar 0.8264 yang didapat dari alat bantu *Rapidminer*.
4. Penelitian ini dilakukan oleh [28] dengan judul “Perbandingan Algoritma *Machine Learning* pada Klasifikasi Penyakit Jantung”. Penelitian dilakukan untuk mengklasifikasi penyakit jantung dengan menggunakan metode *Logistic Regression*, *Naive Bayes*, dan K-Nearest Neighbour (K-NN), dan hasil terbaik pada penelitian ini diperoleh dengan metode *K-Nearest Neighbor* (K-NN) dengan akurasi sebesar 91%, sedangkan akurasi pada algoritma *Logistic Regression* sebesar 85% dan akurasi pada metode *Naive Bayes* sebesar 83%.
5. Penelitian ini dilakukan oleh [29] dengan judul “Penerapan Algoritma *K-Nearest Neighbor* Dalam Klasifikasi Status *Stunting* Balita Di Desa Bojongemas”. Penelitian dilakukan untuk mengklasifikasi status *Stunting* balita di Desa Bojongemas dengan menggunakan metode *K-Nearest Neighbor*, dan hasil akurasi sebesar 92% dari penelitian ini dengan jumlah data sebanyak 503 data.

III. METODOLOGI PENELITIAN

A. Proses dan Langkah Penelitian

Didalam bab ini akan menjelaskan tentang Tahapan - tahapan metodologi penelitian yang dapat dilihat pada gambar 1 dibawah ini :



Gbr. 1 Alur Metodologi Penelitian

B. Metode Pengolahan dan Analisis Data

Metode ini adalah bagian dimana penulis akan menjelaskan tahapan – tahapan pengerjaan yang ada didalam penelitian ini dan berikut ini adalah penjelasannya :

1) Pengumpulan Data

Pada tahapan ini data ambil dari *website* resmi (Badan Pusat Statistik, 2024) yaitu data “Rata-rata Lama Sekolah (RLS) menurut Jenis Kelamin (Tahun), 2019-2021 dan 2022-2023” dan data tersebut di *download* dalam bentuk *CSV* atau *Excel*. Data yang berada pada *website* dapat dilihat pada gambar 2 dan 3 dibawah ini :

Gbr. 2 Dataset dari tahun 2019-2021

Sumber : [30]

Gbr. 3 Dataset dari tahun 2022-2023

Sumber : [30]

Total *Dataset* yang dikumpulkan sebanyak 1.028 data dari seluruh kota dan kabupaten di Indonesia, yang memiliki atribut yaitu : Nama Kota dan Kabupaten, Tahun dan Jenis Kelamin.

2) Labeling

Pada tahapan *Dataset* yang telah didapatkan diberikan *label* atau tanda untuk memudahkan proses klasifikasi, didalam penelitian ini pelabelan dilakukan dengan mencari rata-rata dari *Dataset* tersebut kemudian dibagi 2 untuk menentukan *label* LAMA dan SEBENTAR, yang dapat dilihat pada gambar 4 dibawah ini :

kota dan Kabupaten (P) 19-21	2019	2020	2021	RATA-RATA	LABEL
1 Aceh Barat	8,84	9,15	9,39	9,126667	LAMA
2 Aceh Barat Daya	8,3	8,52	8,53	8,45	LAMA
3 Aceh Besar	10,1	10,11	10,13	10,113333	LAMA
4 Aceh Jaya	8,41	8,59	8,6	8,533333	LAMA
5 Aceh Selatan	8,1	8,41	8,42	8,31	LAMA
6 Aceh Singkil	7,94	7,95	8,17	8,02	LAMA
7 Aceh Tamiang	8,66	8,67	8,83	8,72	LAMA
8 Aceh Tengah	9,64	9,68	9,69	9,67	LAMA
9 Aceh Tenggara	9,12	9,13	9,21	9,153333	LAMA
10 Aceh Timur	7,51	7,84	7,96	7,77	LAMA
11 Aceh Utara	7,97	8,19	8,31	8,156667	LAMA
12 Agam	8,82	9,05	9,06	8,976667	LAMA
13 Alor	7,68	8	8,01	7,896667	LAMA
14 Asahan	8,26	8,61	8,62	8,496667	LAMA
15 Asmat	4,05	4,29	4,43	4,256667	SEBENTAR
16 Badung	9,69	9,7	9,97	9,786667	LAMA
17 Balangan	6,71	7,04	7,19	6,98	LAMA
18 Bandung	8,49	8,71	8,82	8,673333	LAMA
19 Bandung Barat	7,81	7,82	7,83	7,82	LAMA
20 Banggai	7,82	8,12	8,13	8,023333	LAMA
21 Banggai Kepulauan	7,99	8,25	8,26	8,166667	LAMA
22 Banggai Laut	8,01	8,15	8,16	8,106667	LAMA
23 Bangka	7,98	7,99	8	7,99	LAMA
24 Bangka Barat	7,18	7,19	7,39	7,253333	LAMA
25 Bangka Selatan	5,98	6,2	6,46	6,213333	LAMA
26 Bangka Tengah	6,73	6,9	6,91	6,846667	LAMA
27 Bangkalan	4,97	5,25	5,35	5,19	SEBENTAR
28 Bangli	6,31	6,47	6,48	6,42	LAMA
29 Banjar	6,81	6,82	7,08	6,903333	LAMA
30 Banjarnegara	6,32	6,57	6,58	6,49	LAMA
31 Bantaeng	6,41	6,69	6,74	6,613333	LAMA
32 Bantul	9,04	9,05	9,28	9,123333	LAMA
33 Banyu Asin	6,87	6,88	7,09	6,946667	LAMA
34 Banyumas	7,14	7,24	7,35	7,243333	LAMA
35 Banyuwangi	6,54	6,55	6,79	6,626667	SEBENTAR
36	6,75	6,75	6,75	6,75	LAMA

Gbr. 4 Labeling

Seperti pada gambar 4 data dari tahun 2019 – 2021 dicari rata-ratanya kemudian mencari rata-rata paling tinggi dan dibagikan 2 setelah itu, rata-rata yang menengah kebawah dikategorikan sebagai SEBENTAR dan menengah keatas dikategorikan sebagai LAMA.

3) Data Table

Tahapan ini adalah tahapan *Pre-Processing* yaitu *Data Table* tahapan untuk mengolah kumpulan data dan dikeluarkan menjadi *spreadsheet* sehingga dapat diproses oleh algoritma.

4) Select Column

Tahapan kedua dari *Pre-Processing* adalah *Select Column*, pada tahapan ini atribut yang ada didata, dipilih secara manual dan dikelompokkan menjadi 4 bagian yaitu : *Ignored* (Atribut

yang tidak diinginkan), *Features* (Atribut yang berguna untuk penelitian), *Target* (Atribut yang akan menjadi tujuan dari pengolahan data) dan *Meta* (Atribut pelengkap yang bisa mempengaruhi dan bisa juga tidak mempengaruhi pengolahan data).

5) *K-Nearest Neighbor (K-NN)*

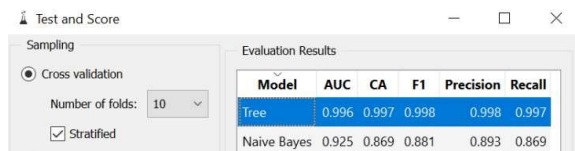
Pada tahap ini peneliti memakai Algoritma *K-Nearest Neighbor (K-NN)* untuk mengolah *Dataset* yang ada, karena K-NN dapat mengurangi noise data dan mengolah data dalam jumlah besar.

6) *Cross Validation*

Pada tahapan ini data yang diproses algoritma dibagi menjadi dua kategori yaitu menjadi *data training* (data latih) dan *data testing* (data uji) dan dilakukan secara berulang sampai setiap data mendapatkan porsi yang sama dalam *training* dan *testing* kemudian hasilnya akan ditampilkan dalam bentuk statistik.

7) *Test and Score*

Pada bagian ini semua data yang telah diolah dikeluarkan kedalam bentuk statistik yang dibagi jadi 5 bagian yang dapat dilihat pada gambar 5 dan penjelasannya ada dibawah gambar tersebut.



Gbr. 5 Statistik dari *Test and Score*

Sumber : [15]

Penjelasannya :

1. *AUC* : *AUC (Area Under Curve)* adalah hasil dari kinerja pengolahan data, dari *AUC* ini kita bisa menentukan apakah hasil pengolahan data dari algoritma itu hasilnya bagus.
2. *CA* : *CA (Classification Accuracy)* menunjukkan berapa banyak prediksi benar (positif dan negatif) dari seluruh data yang diolah.
3. *F1* : *F1* menunjukkan hasil akurasi dari evaluasi pengolahan data dengan menggabungkan presisi (*Precision*) dan perolehan (*Recall*).
4. *Precision* : *Precision* menunjukkan perbandingan antara TP dan FP dimana semakin banyak FP maka hasil *Precision* akan semakin rendah.
5. *Recall* : *Recall* menunjukkan perbandingan antara TP dan FN dimana semakin banyak FN maka hasil *Recall* akan semakin rendah.

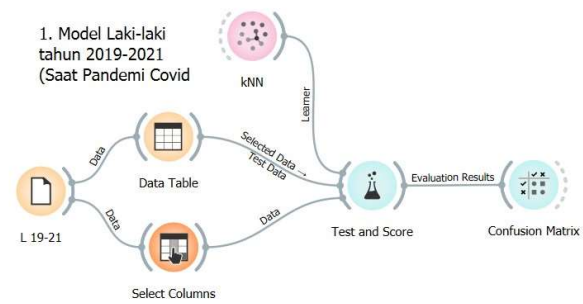
8) *Confusion Matrix*

Pada tahapan ini *Confusion Matrix* digunakan untuk melihat hasil pengolahan data, dimana hasil tersebut ditampilkan dalam bentuk tabel yang berisikan nilai dari hasil tersebut yaitu TN, FP, FN, TN.

9) *Orange Data mining*

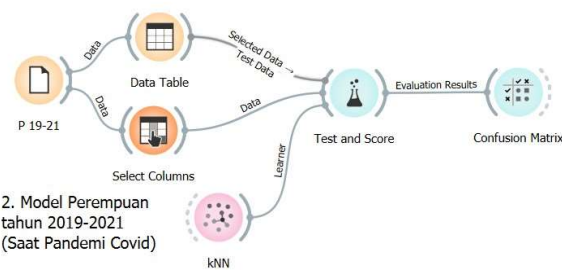
*Orange Data mining* adalah *software* yang digunakan untuk mengolah data dengan bantuan algoritma *Machine Learning*. Didalam penelitian ini peneliti membuat 4 model menggunakan *Orange Data mining* yaitu :

- Model 1 : Untuk mengolah data laki-laki dari tahun 2019-2021. Yang dapat dilihat pada gambar 6 dibawah ini.



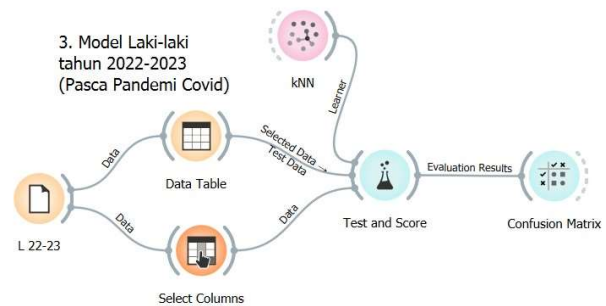
Gbr. 6 Model laki-laki dari tahun 2019-2021

- Model 2 : Untuk mengolah data perempuan dari tahun 2019-2021. Yang dapat dilihat pada gambar 7 dibawah ini.



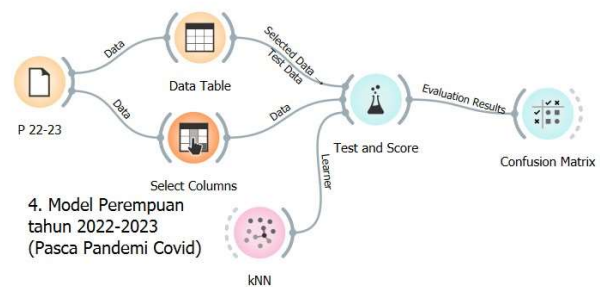
Gbr. 7 Model perempuan dari tahun 2019-2021

- Model 3 : Untuk mengolah data laki-laki dari tahun 2022-2023. Yang dapat dilihat pada gambar 8 dibawah ini.



Gbr. 8 Model laki-laki dari tahun 2022-2023

- Model 4 : Untuk mengolah data perempuan dari tahun 2022-2023. Yang dapat dilihat pada gambar 9 dibawah ini.



Gbr. 9 Model perempuan dari tahun 2022-2023



IV. HASIL DAN PEMBAHASAN

A. Hasil Penelitian

1) Hasil Data Collections

Didalam penelitian ini, *Dataset* yang diambil yaitu 1.028 data, dan data tersebut dibagi menjadi 4 bagian kemudian diberikan *label* LAMA (Positif) dan SEBENTAR (Negatif) yang rinciannya dapat dilihat pada tabel II, dan dokumentasinya dapat dilihat pada gambar 10, 11, 12, dan 13 dibawah ini.

Tabel II  
Rincian *Dataset*

No.	Data	Tahun	Banyak Data
1	Laki-laki	2019-2021	514
2	Perempuan	2019-2021	514
3	Laki-laki	2022-2023	514
4	Perempuan	2022-2023	514

1	Kota dan Kabupaten (L) 19-21	2019	2020	2021	RATA-RATA	LABEL
2	Aceh Barat	9,46	9,71	9,82	9,663333	LAMA
3	Aceh Barat Daya	8,59	8,92	8,93	8,813333	LAMA
4	Aceh Besar	10,51	10,52	10,53	10,52	LAMA
5	Aceh Jaya	8,88	8,89	8,9	8,89	LAMA
6	Aceh Selatan	9,07	9,19	9,2	9,153333	LAMA
7	Aceh Singkil	8,94	8,96	9,01	8,97	LAMA
8	Aceh Tamiang	9,13	9,14	9,15	9,14	LAMA
9	Aceh Tengah	9,85	10,02	10,03	9,966667	LAMA
10	Aceh Tenggara	10,2	10,21	10,22	10,21	LAMA
11	Aceh Timur	8,3	8,56	8,57	8,476667	LAMA
12	Aceh Utara	8,83	8,86	8,87	8,853333	LAMA
13	Agam	8,88	8,89	8,9	8,89	LAMA
14	Alor	8,56	8,88	8,89	8,776667	LAMA
15	Asahan	8,83	8,94	8,95	8,906667	LAMA
16	Asmat	5,76	5,86	6,04	5,886667	SEBENTAR
17	Badung	10,92	10,93	11,08	10,97667	LAMA
18	Balangan	8,04	8,28	8,4	8,24	LAMA
19	Bandung	9	9,13	9,3	9,143333	LAMA
20	Bandung Barat	8,7	8,74	8,77	8,74	LAMA

Gbr. 10 Laki-laki 2019-2021

1	Kota dan Kabupaten (P) 19-21	2019	2020	2021	RATA-RATA	LABEL
2	Aceh Barat	8,84	9,15	9,39	9,126667	LAMA
3	Aceh Barat Daya	8,3	8,52	8,53	8,45	LAMA
4	Aceh Besar	10,1	10,11	10,13	10,113333	LAMA
5	Aceh Jaya	8,41	8,59	8,6	8,533333	LAMA
6	Aceh Selatan	8,1	8,41	8,42	8,31	LAMA
7	Aceh Singkil	7,94	7,95	8,17	8,02	LAMA
8	Aceh Tamiang	8,66	8,67	8,83	8,72	LAMA
9	Aceh Tengah	9,64	9,68	9,69	9,67	LAMA
10	Aceh Tenggara	9,12	9,13	9,21	9,153333	LAMA
11	Aceh Timur	7,51	7,84	7,96	7,77	LAMA
12	Aceh Utara	7,97	8,19	8,31	8,156667	LAMA
13	Agam	8,82	9,05	9,06	8,976667	LAMA
14	Alor	7,68	8	8,01	7,896667	LAMA
15	Asahan	8,26	8,61	8,62	8,496667	LAMA
16	Asmat	4,05	4,29	4,43	4,256667	SEBENTAR
17	Badung	9,69	9,7	9,97	9,786667	LAMA
18	Balangan	6,71	7,04	7,19	6,98	LAMA
19	Bandung	8,49	8,71	8,82	8,673333	LAMA
20	Bandung Barat	7,91	7,97	7,97	7,97	LAMA

Gbr. 11 Perempuan 2019-2021

1	Kota dan Kabupaten (L) 22-23	2022	2023	RATA-RATA	LABEL
2	Aceh Barat	10,15	10,16	10,155	LAMA
3	Aceh Barat Daya	8,94	8,96	8,95	LAMA
4	Aceh Besar	10,55	10,56	10,555	LAMA
5	Aceh Jaya	8,92	8,93	8,925	LAMA
6	Aceh Selatan	9,21	9,23	9,22	LAMA
7	Aceh Singkil	9,02	9,03	9,025	LAMA
8	Aceh Tamiang	9,28	9,51	9,395	LAMA
9	Aceh Tengah	10,04	10,06	10,05	LAMA
10	Aceh Tenggara	10,24	10,39	10,315	LAMA
11	Aceh Timur	8,58	8,6	8,59	LAMA
12	Aceh Utara	9,02	9,03	9,025	LAMA
13	Agam	8,92	9,09	9,005	LAMA
14	Alor	8,9	8,92	8,91	LAMA
15	Asahan	8,97	8,99	8,98	LAMA
16	Asmat	6,36	6,6	6,48	SEBENTAR
17	Badung	11,09	11,32	11,205	LAMA
18	Balangan	8,64	8,78	8,71	LAMA
19	Bandung	9,31	9,32	9,315	LAMA
20	Bandung Barat	8,74	8,75	8,745	LAMA

Gbr. 12 Laki-laki 2022-2023

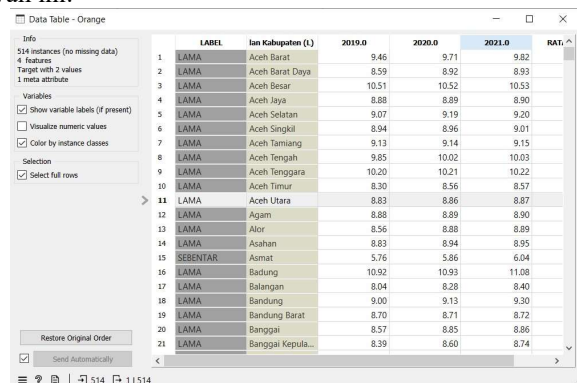
1	Kota dan Kabupaten (P) 22-23	2022	2023	RATA-RATA	LABEL
2	Aceh Barat	9,58	9,8	9,69	LAMA
3	Aceh Barat Daya	8,54	8,66	8,6	LAMA
4	Aceh Besar	10,31	10,36	10,335	LAMA
5	Aceh Jaya	8,61	8,63	8,62	LAMA
6	Aceh Selatan	8,43	8,45	8,44	LAMA
7	Aceh Singkil	8,21	8,28	8,245	LAMA
8	Aceh Tamiang	8,84	9,01	8,925	LAMA
9	Aceh Tengah	9,7	9,72	9,71	LAMA
10	Aceh Tenggara	9,55	9,81	9,68	LAMA
11	Aceh Timur	8,13	8,38	8,255	LAMA
12	Aceh Utara	8,4	8,72	8,56	LAMA
13	Agam	9,07	9,33	9,2	LAMA
14	Alor	8,02	8,04	8,03	LAMA
15	Asahan	8,64	8,65	8,645	LAMA
16	Asmat	4,7	4,83	4,765	SEBENTAR
17	Badung	10,07	10,52	10,295	LAMA
18	Balangan	7,35	7,61	7,48	LAMA
19	Bandung	8,83	8,93	8,88	LAMA
20	Bandung Barat	7,95	7,96	7,955	LAMA

Gbr. 13 Perempuan 2022-2023

2) Hasil Pre-Processing

- Data Table

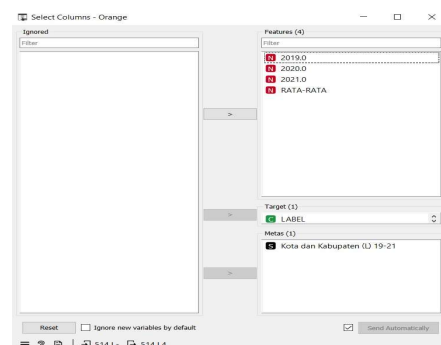
Hasil dari *Data Table* pada penelitian ini adalah *Dataset* berupa *Excel* diproses menggunakan *Data Table* kemudian dikeluarkan dalam bentuk *Spreadsheet* yang lebih terstruktur sehingga mudah diolah oleh *Algoritma Machine Learning*, dokumentasi dari hasil *Data Table* ada pada gambar 14 dibawah ini.



Gbr. 14 Hasil *Data Table*

- Select Column

Hasil dari *Select Column* adalah *Dataset* yang memiliki atribut, dibagi menjadi 3 melalui *Select Column* dengan kategori *Features*, *Target* dan *Meta*. Proses pengkategorian dilakukan secara manual. *Features* itu adalah atribut yang digunakan *system* untuk belajar. *Target* adalah tujuan dari pengolahan ini dimana hasil dari pembelajaran dan pelatihan tujuannya adalah *Target*. *Meta* adalah atribut pelengkap. Dokumentasi dari *Select Column* ada pada gambar 15 dibawah ini.



Gbr. 15 Hasil *Select Column*

B. Hasil Pengujian

Metode ini adalah bagian dimana penulis akan menjelaskan tahapan – tahapan pengerjaan yang ada didalam penelitian ini dan berikut ini adalah penjelasannya :

1) Hasil K-Nearest Neighbor (K-NN) Model 1

- Cross Validation Model 1

Data yang diolah pada tahap Cross Validation melalui tahap 5, 10, dan 20 Folds. Folds adalah proses mengacak data sehingga data tersebut mendapatkan training dan testing yang sama dan hasilnya dapat dilihat pada tabel III dibawah ini.

Tabel III  
Cross Validation Model 1

Folds	AUC	CA	F1	Precision	Recall
5	0.908	0.984	0.983	0.984	0.984
10	0.909	0.984	0.983	0.984	0.984
20	0.909	0.984	0.983	0.984	0.984

Dari data diatas terlihat ada perubahan AUC dari 5 Folds ke 10 Folds, kemudian dari 10 Folds ke 20 Folds tidak ada perubahan dari hasil data tersebut.

- Confusion Matrix Model 1

Pada Confusion Matrix ini peneliti memakai 20 Folds karena hasilnya menunjukkan yang tertinggi, sehingga hasil dapat dilihat pada gambar 16.

		Predicted		Σ
		LAMA	SEBENTAR	
Actual	LAMA	490	1	491
	SEBENTAR	7	16	23
Σ		497	17	514

Gbr. 16 Confusion Matrix Model 1

2) Hasil K-Nearest Neighbor (K-NN) Model 2

- Cross Validation Model 2

Data yang diolah pada tahap Cross Validation melalui tahap 5, 10, dan 20 Folds. Folds adalah proses mengacak data sehingga data tersebut mendapatkan training dan testing yang sama dan hasilnya dapat dilihat pada tabel IV dibawah ini.

Tabel IV  
Cross Validation Model 2

Folds	AUC	CA	F1	Precision	Recall
5	0.823	0.969	0.965	0.970	0.969
10	0.823	0.967	0.963	0.967	0.967
20	0.848	0.963	0.959	0.962	0.963

Dari data diatas terlihat tidak ada perubahan AUC dari 5 Folds ke 10 Folds, kemudian dari 10 Folds ke 20 Folds ada perubahan dari hasil data tersebut.

- Confusion Matrix Model 2

Pada Confusion Matrix ini peneliti memakai 20 Folds karena hasilnya menunjukkan yang tertinggi, sehingga hasil dapat dilihat pada gambar 17.

		Predicted		Σ
		LAMA	SEBENTAR	
Actual	LAMA	474	1	475
	SEBENTAR	16	23	39
Σ		490	24	514

Gbr. 17 Confusion Matrix Model 2

3) Hasil K-Nearest Neighbor (K-NN) Model 3

- Cross Validation Model 3

Data yang diolah pada tahap Cross Validation melalui tahap 5, 10, dan 20 Folds. Folds adalah proses mengacak data sehingga data tersebut mendapatkan training dan testing yang sama dan hasilnya dapat dilihat pada tabel V dibawah ini.

Tabel V  
Cross Validation Model 3

Folds	AUC	CA	F1	Precision	Recall
5	0.845	0.979	0.977	0.977	0.979
10	0.846	0.977	0.975	0.975	0.977
20	0.846	0.977	0.975	0.975	0.977

Dari data diatas terlihat ada perubahan AUC dari 5 Folds ke 10 Folds, kemudian dari 10 Folds ke 20 Folds tidak ada perubahan dari hasil data tersebut.

- Confusion Matrix Model 3

Pada Confusion Matrix ini peneliti memakai 20 Folds karena hasilnya menunjukkan yang tertinggi, sehingga hasil dapat dilihat pada gambar 18.

		Predicted		Σ
		LAMA	SEBENTAR	
Actual	LAMA	490	3	493
	SEBENTAR	9	12	21
Σ		499	15	514

Gbr. 18 Confusion Matrix Model 3

4) Hasil K-Nearest Neighbor (K-NN) Model 4

- Cross Validation Model 4

Data yang diolah pada tahap Cross Validation melalui tahap 5, 10, dan 20 Folds. Folds adalah proses mengacak data sehingga data tersebut mendapatkan training dan testing yang sama dan hasilnya dapat dilihat pada tabel VI dibawah ini.

Tabel VI  
Cross Validation Model 4

Folds	AUC	CA	F1	Precision	Recall
5	0.919	0.981	0.980	0.980	0.981
10	0.919	0.981	0.980	0.980	0.981
20	0.920	0.981	0.980	0.980	0.981

Dari data diatas terlihat tidak ada perubahan AUC dari 5 Folds ke 10 Folds, kemudian dari 10 Folds ke 20 Folds ada perubahan dari hasil data tersebut.

- Confusion Matrix Model 4

Pada *Confusion Matrix* ini peneliti memakai 20 *Folds* karena hasilnya menunjukkan yang tertinggi, sehingga hasil dapat dilihat pada gambar 19.

		Predicted		$\Sigma$
		LAMA	SEBENTAR	
Actual	LAMA	463	3	466
	SEBENTAR	7	41	48
$\Sigma$		470	44	514

Gbr. 19 Confusion Matrix Model 4

5) Hipotesis

Dari hasil penelitian dan pengujian di atas, hipotesis H1 terbukti benar, yaitu adanya hasil yang signifikan dalam mengklasifikasi data masa studi di Indonesia berdasarkan jenis kelamin. Sehingga penelitian ini dapat menjadi referensi bagi pemerintah Indonesia agar dapat meningkatkan kualitas pendidikan di Indonesia. Dengan hasil yang signifikan ini, diharapkan pemerintah Indonesia dapat mengambil langkah-langkah strategis untuk mengurangi kesenjangan *gender* dalam pendidikan dan menciptakan kesempatan belajar yang lebih merata bagi seluruh lapisan masyarakat.

Dengan demikian, hasil penelitian ini juga diharapkan dapat mendorong peninjauan ulang kebijakan pendidikan yang ada, serta menjadi dasar dalam penyusunan program-program pendidikan yang lebih inklusif dan responsif terhadap perbedaan *gender*. Implementasi kebijakan yang berfokus pada peningkatan akses dan kualitas pendidikan bagi semua *gender* akan sangat membantu dalam menciptakan generasi yang lebih kompetitif dan berdaya saing tinggi di kancah global.

V. KESIMPULAN

Dari hasil penelitian dan pengujian yang telah penulis lakukan sehingga didapatkan kesimpulan sebagai berikut :

Hasil dari pengolahan data menggunakan algoritma *Machine Learning* memberikan hasil yang optimal terbukti dari data sebanyak 1.028 data yang diambil dari Badan Pusat Statistik meliputi data dari tahun 2019-2023, serta dilakukan pelabelan pada data tersebut dengan cara mencari rata-rata dan di lakukan pelabelan Lama dan Sebentar, memberikan hasil akurasi terhadap 4 model menggunakan 20 *Folds* yaitu [0.909], [0.848], [0.846], dan [0.920].

Pengolahan data dengan bantuan algoritma *Machine Learning* mudah untuk dipahami dan direalisasikan karena setelah mendapatkan hasil pengujian, didapatkan bahwa Saat Pandemi *Covid* Laki-laki mendapatkan pendidikan yang Lama dan Pasca Pandemi *Covid* Laki-laki juga mendapatkan pendidikan yang Lama, dan masalah jenis kelamin juga terbukti karena Perempuan mendapatkan pendidikan tidak se lama Laki-laki.

Pemerintah dapat lebih memperhatikan lagi kualitas pendidikan di Indonesia, terlebih Laki-laki dan Perempuan harus mendapatkan pendidikan yang setara sehingga Sumber Daya Manusia yang dihasilkan akan berkualitas juga dan itu yang akan menjadikan Negara Indonesia ini lebih maju lagi, serta daerah-daerah yang minim perhatian seperti dibagian Timur Indonesia, pendidikannya harus di tingkatkan.

Hasil penelitian dan pengujian di atas, hipotesis H1 terbukti benar, yaitu adanya hasil yang signifikan dalam mengklasifikasi data masa studi di Indonesia berdasarkan jenis kelamin. Sehingga penelitian ini dapat menjadi referensi bagi pemerintah Indonesia agar dapat meningkatkan kualitas pendidikan di Indonesia. Dengan hasil yang signifikan ini, diharapkan pemerintah Indonesia dapat mengambil langkah-langkah strategis untuk mengurangi kesenjangan *gender* dalam pendidikan dan menciptakan kesempatan belajar yang lebih merata bagi seluruh lapisan masyarakat.

REFERENSI

- [1] A. O. Safitri, V. D. Yuniarti, and D. Rostika, "Upaya Peningkatan Pendidikan Berkualitas di Indonesia: Analisis Pencapaian Sustainable Development Goals (SDGs)," *Jurnal Basicedu*, vol. 6, no. 4, pp. 7096–7106, Jun. 2022, doi: 10.31004/basicedu.v6i4.3296.
- [2] H. Yuslin and A. Irfan, "KESETARAAN GENDER DI BIDANG PENDIDIKAN DAN PERTUMBUHAN EKONOMI DI PROVINSI SULAWESI SELATAN," *AN-NISA*, vol. 15, no. 1, pp. 29–37, Jun. 2022, doi: 10.30863/an.v15i1.3545.
- [3] V. A. Saputra and S. A. Amomo, "PENERAPAN METODE MACHINE LEARNING DALAM MENGIDENTIFIKASI BERITA HOAKS," *Computer Based Information System Journal*, vol. 12, no. 1, pp. 112–121, Mar. 2024, doi: 10.33884/cbis.v12i1.8442.
- [4] R. Kurniawan, P. B. Wintoro, Y. Mulyani, and M. Komarudin, "IMPLEMENTASI ARSITEKTUR XCEPTION PADA MODEL MACHINE LEARNING KLASIFIKASI SAMPAH ANORGANIK," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 11, no. 2, Apr. 2023, doi: 10.23960/jitet.v11i2.3034.
- [5] M. F. Muttaqin, T. Bukhori, Y. Yanto, N. Agustina, and M. Naseer, "SISTEM PREDIKSI BERITA PALSU TENTANG VIRUS COVID-19 MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE (SVM)," *Naratif: Jurnal Nasional Riset, Aplikasi dan Teknik Informatika*, vol. 5, no. 1, pp. 26–33, Jun. 2023, doi: 10.53580/naratif.v5i1.187.
- [6] B. R. Setiaji, D. Q. Utama, and A. Adiwijaya, "Smartphone Purchase Recommendation System Using the K-Nearest Neighbor (KNN) Algorithm," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 6, no. 4, p. 2180, Oct. 2022, doi: 10.30865/mib.v6i4.4753.
- [7] K. Shahnazari and S. M. Ayyoubzadeh, "A Novel Nearest Neighbors Algorithm Based on Power Muirhead Mean," *Cornell University*, Sep. 2022.
- [8] M. Rijal, Adnan, and A. Achmad, "KLASIFIKASI PENYAKIT PERNAPASAN BERBASIS VISUALISASI SUARA MENGGUNAKAN METODE SUPPORT VECTOR MACHINE," *JURNAL ILMIAH ILMU KOMPUTER*, vol. 8, no. 2, pp. 115–119, Oct. 2022, doi: 10.35329/jiik.v8i2.245.
- [9] Pande Made Risky Cahya Dinatha and Nur Aini Rakhmawati, "Komparasi Term Weighting dan Word Embedding pada Klasifikasi Tweet Pemerintah Daerah," *Jurnal Nasional Teknik Elektro dan Teknologi Informasi*, vol. 9, no. 2, pp. 155–161, May 2020, doi: 10.22146/jnteti.v9i2.90.
- [10] F. Siva, S. M. U. Assegaf, S. A. Pahlevi, and M. A. Yaqin, "Survei Metode-Metode Software Development Life Cycle dengan Metode Systematic Literature Review," *ILKOMNIKA: Journal of Computer Science and Applied Informatics*, vol. 5, no. 2, pp. 36–52, Aug. 2023, doi: 10.28926/ilkomnika.v5i2.447.
- [11] S. S. Rahmatia, I. Darmawan, A. S. Dewi, and S. N. Khasanah, "Analisis Kesuksesan Aplikasi Alputak Betawi Menggunakan Metode Delone And McLean," *Jurnal Ilmiah FIFO*, vol. 15, no. 1, p. 67, Feb. 2024, doi: 10.22441/fifo.2023.v15i1.008.
- [12] S. S. Salmia, "DEVELOPMENT OF QUALITY INSTRUMENTS AND DATA COLLECTION TECHNIQUES," *Jurnal Pendidikan dan Pengajaran Guru Sekolah Dasar (JPPGuseda)*, vol. 6, no. 1, pp. 119–124, Mar. 2023, doi: 10.55215/jppguseda.v6i1.7527.
- [13] D. F. Zahra, "Analisis Sentimen penggunaan MyPertamina untuk Pembelian BBM Bersubsidi menggunakan Algoritma Naive Bayes," *Explore: Jurnal Sistem Informasi dan Telematika*, vol. 14, no. 1, p. 94, Jul. 2023, doi: 10.36448/jsit.v14i1.3098.
- [14] Y. Zhang, Y. Wang, H. Zhang, B. Zhu, S. Chen, and D. Zhang, "OneLabeler: A Flexible System for Building Data Labeling Tools," in *CHI Conference on Human Factors in Computing Systems*, New



- York, NY, USA: ACM, Apr. 2022, pp. 1–22. doi: 10.1145/3491102.3517612.
- [15] L. Irawan, L. H. Hasibuan, and F. Fauzi, “ANALISA PREDIKSI EFEK KERUSAKAN GEMPA DARI MAGNITUDO (SKALA RICHTER) DENGAN METODE ALGORITMA ID3 MENGGUNAKAN APLIKASI DATA MINING ORANGE,” *Jurnal Teknologi Informasi: Jurnal Keilmuan dan Aplikasi Bidang Teknik Informatika*, vol. 14, no. 2, pp. 189–201, Aug. 2020, doi: 10.47111/jti.v14i2.1079.
- [16] A. C. Handoko and H. Hendry, “PERBANDINGAN METODE SUPERVISED LEARNING UNTUK PREDIKSI DIABETES GESTASIONAL DENGAN SOFTWARE ORANGE,” *JUPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 8, no. 4, pp. 1238–1247, Nov. 2023, doi: 10.29100/jupi.v8i4.4166.
- [17] Aldi Prajela, Syafrandi Syafrandi, Dony Permana, and Dina Fitria, “Twitter Data Sentimen Analysis 2024 Presidential Candidate Using Algorithm Naïve Bayes Classifier By Methods K-Fold Cross Validation,” *UNP Journal of Statistics and Data Science*, vol. 2, no. 1, pp. 99–104, Feb. 2024, doi: 10.24036/ujsds/vol2-iss1/149.
- [18] P. Adytia, W. Wahyuni, K. Sussolaikah, and Y. Satria, “KLASIFIKASI PENGGUNAAN DATA TRAFIK INTERNET MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE,” *Jurnal Komputer dan Informatika*, vol. 11, no. 1, pp. 96–102, Apr. 2023, doi: 10.35508/jicon.v11i1.10039.
- [19] A. A. Bagaskara and K. D. Hartomo, “Classification of Flood-Prone Areas Using 10-Fold Cross Validation and K-Nearest Neighbors,” *SISTEMASI*, vol. 13, no. 1, p. 315, Jan. 2024, doi: 10.32520/stmsi.v13i1.3637.
- [20] G. C. Triasis, D. Arisandi, and T. Sutrisno, “ANALISIS KEPUASAN PENGGUNAAN APLIKASI SHOPEE MENGGUNAKAN ALGORITMA NAÏVE BAYES,” *Jurnal Ilmu Komputer dan Sistem Informasi*, vol. 10, no. 1, Mar. 2022, doi: 10.24912/jiksi.v10i1.17857.
- [21] N. I. Nurhidayati, Y. Yahya, F. Fathurrahman, L. M. Samsu, and W. Amnia, “Implementasi Algoritma Naive Bayes Untuk Klasifikasi Penerima Beasiswa (Studi Kasus Universitas Hamzanwadi),” *Infotek : Jurnal Informatika dan Teknologi*, vol. 6, no. 1, pp. 177–188, Jan. 2023, doi: 10.29408/jit.v6i1.7529.
- [22] Ratnasari, A. Jurnaidi Wahidin, A. Eko Setiawan, and P. Bintoro, “Machine Learning Untuk Klasifikasi Penyakit Jantung,” *Aisyah Journal Of Informatics and Electrical Engineering (A.J.I.E.E)*, vol. 6, no. 1, pp. 145–150, Feb. 2024, doi: 10.30604/jti.v6i1.272.
- [23] M. A. J. Al-Sammaraie et al., “Predicting Fruit’s Sweetness Using Artificial Intelligence—Case Study: Orange,” *Applied Sciences*, vol. 12, no. 16, p. 8233, Aug. 2022, doi: 10.3390/app12168233.
- [24] A. Arabiat and M. Altayeb, “Assessing the effectiveness of data mining tools in classifying and predicting road traffic congestion,” *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 34, no. 2, p. 1295, May 2024, doi: 10.11591/ijeecs.v34.i2.pp1295-1303.
- [25] K. H. L. Malelak, I. M. D. Ardiada, and G. Feoh, “IMPLEMENTASI KLASIFIKASI NAIVE BAYES DALAM MEMPREDIKSI LAMA STUDI MAHASISWA (STUDI KASUS: UNIVERSITAS DHYANA PURA),” *SINTECH (Science and Information Technology) Journal*, vol. 4, no. 2, pp. 202–209, Oct. 2021, doi: 10.31598/sintechjournal.v4i2.964.
- [26] D. Safitri, S. S. Hilabi, and F. Nurapriani, “ANALISIS PENGGUNAAN ALGORITMA KLASIFIKASI DALAM PREDIKSI KELULUSAN MENGGUNAKAN ORANGE DATA MINING,” *Rabit : Jurnal Teknologi dan Sistem Informasi Univrab*, vol. 8, no. 1, pp. 75–81, Jan. 2023, doi: 10.36341/rabit.v8i1.3009.
- [27] M. R. Qisthiano, “PENERAPAN MODEL KLASIFIKASI NAÏVE BAYES UNTUK PREDIKSI MAHASISWA TEPAT WAKTU,” *Semnas Ristek (Seminar Nasional Riset dan Inovasi Teknologi)*, vol. 7, no. 1, Jan. 2023, doi: 10.30998/semnasristek.v7i1.6266.
- [28] S. Heristian and M. Napiah, “Perbandingan Algoritma Machine Learning pada Klasifikasi Penyakit Jantung,” *Jurnal Infortech*, vol. 6, no. 1, pp. 46–51, Jun. 2024, doi: 10.31294/infortech.v6i1.21888.
- [29] S. Widia Pebrianti, R. Astuti, and F. M. Basysyar, “PENERAPAN ALGORITMA K-NEAREST NEIGHBOR DALAM KLASIFIKASI STATUS STUNTING BALITA DI DESA BOJONGEMAS,” *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 2, pp. 2479–2488, Apr. 2024, doi: 10.36040/jati.v8i2.8448.
- [30] Badan Pusat Statistik, “Rata-rata Lama Sekolah (RLS) menurut Jenis Kelamin (Tahun), 2019-2023,” *bps.go.id*. Accessed: Jun. 19, 2024. [Online]. Available: <https://www.bps.go.id/id/statistics-table/2/NDU5IzI=/rata-rata-lama-sekolah-rls-menurut-jenis-kelamin.html>