

Penerapan Metode Naive Bayes Untuk Penentuan Bibit Kelapa Sawit Berdasarkan Kondisi Daerah Tanam dan Perawatan Tanaman

Ansfridus Bravo¹, Tursina², Helen Sastypratiwi³

*Program Studi Informatika, Fakultas Teknik, Universitas Tanjungpura
Jl. Prof. Dr. H. Hadari Nawawi, Pontianak 78124*

¹bravobb@student.untan.ac.id

²tursina@informatika.untan.ac.id

³helensastypratiwi@informatics.untan.ac.id

Abstrak

Pemilihan bibit kelapa sawit yang tepat merupakan salah satu upaya dalam meningkatkan produktifitas tanaman kelapa sawit. Pada saat pemilihan bibit, permasalahan yang sering dihadapi adalah tidak semua jenis bibit dapat sesuai terhadap kondisi lahan dan jenis perawatan yang berbeda-beda. Oleh karena itu, dibutuhkan sebuah sistem yang membantu dalam menentukan jenis bibit yang sesuai dengan kondisi lahan dan perawatan yang diberikan. Dalam penelitian ini penulis menerapkan model prediksi machine learning dalam penentuan bibit kelapa sawit berdasarkan kondisi lahan dan perawatan tanaman. Tujuan penelitian ini adalah menghasilkan model klasifikasi penentuan bibit kelapa sawit berdasarkan kondisi daerah tanam dan perawatan. Metode yang digunakan dalam penelitian ini adalah naïve bayes dengan dua jenis naïve bayes yang berbeda yaitu gaussian naïve bayes dan multinomial naïve bayes dengan dua skenario pengujian. Hasil evaluasi menunjukkan bahwa performa dari model gaussian naïve bayes dan multinomial naïve pada skenario pengujian pertama bayes dengan nilai accuracy masing-masing sebesar 55% dan 62% serta nilai f1-score masing-masing 63% dan 66%. Sedangkan pada skenario pengujian kedua hasil evaluasi accuracy untuk model gaussian naïve bayes dan multinomial naïve bayes dengan nilai accuracy masing-masing sebesar 70% dan 68% serta f1-score masing-masing sebesar 69% dan 64%. Setelah dilakukan evaluasi model, maka model naïve bayes yang memiliki performa terbaik yaitu gaussian naïve bayes pada skenario pengujian kedua yang digunakan sebagai model prediksi machine learning untuk penentuan bibit kelapa sawit berdasarkan kondisi daerah tanam dan perawatan pada aplikasi generik yang dibangun pada penelitian ini.

Kata kunci: pemilihan bibit, naïve bayes, machine learning

Application of Naive Bayes Method for Determination of Oil Palm Seeds Based on Conditions of Planting Areas and Plant Care

Abstract

Selection of the right oil palm seeds is one of the efforts to increase the productivity of oil palm plants. When selecting seeds, the problem that is often faced is that not all types of seeds can be adapted to different land conditions and types of care. Therefore, we need a system that helps in determining the type of seed that is in accordance with the conditions of the land and the care provided. In this study, the authors apply a machine learning prediction model in determining oil palm seedlings based on land conditions and plant care. The purpose of this study was to produce a classification model for determining oil palm seedlings based on the conditions of the planting and maintenance areas. The method used in this research is naive bayes with two different types of nave bayes, namely gaussian nave bayes and multinomial nave bayes with two test scenarios. The evaluation results show that the performance of the Gaussian naive Bayes and multinomial naive models in the first Bayesian test scenario with an accuracy value of 55% and 62%, respectively, and an f1-score value of 63% and 66%, respectively. While in the second test scenario the accuracy evaluation results for the gaussian naive Bayes and multinomial naive Bayes models with accuracy values of 70% and 68% respectively and f1-scores of 69% and 64%, respectively. After evaluating the model, the naive Bayes model that has the best performance is Gaussian naive Bayes in the second test scenario which is used as a machine learning prediction

model for determining oil palm seedlings based on the conditions of the planting area and maintenance on the generic application built in this study.

Keywords: seed selection, naive bayes, machine learning

I. PENDAHULUAN

Kelapa sawit merupakan tanaman dengan nilai ekonomis yang cukup tinggi karena merupakan salah satu tanaman penghasil minyak nabati. Bagi Indonesia, kelapa sawit memiliki arti penting karena mampu menciptakan kesempatan kerja bagi masyarakat dan sebagai sumber perolehan devisa negara. Kelapa sawit merupakan tanaman daerah tropis yang membutuhkan suhu, jenis tanah, ketinggian daerah dan curah hujan yang cukup untuk mendapatkan hasil yang maksimal [1]. Tanaman Kelapa sawit (*Elaeis guineensis* jacq) saat ini merupakan salah satu jenis tanaman perkebunan yang menduduki posisi penting disektor pertanian umumnya, dan sektor perkebunan khususnya. Hal ini disebabkan karena dari sekian banyak tanaman yang menghasilkan minyak atau lemak, kelapa sawit yang menghasilkan nilai ekonomi terbesar per hektarnya didunia [2].

Melihat pentingnya tanaman kelapa sawit dewasa ini dan masa yang akan datang, seiring dengan meningkatnya kebutuhan penduduk dunia akan minyak sawit, maka perlu dipikirkan usaha peningkatan kualitas dan kuantitas produksi kelapa sawit secara tepat agar sasaran yang diinginkan dapat tercapai. Salah satu diantaranya adalah pemilihan bahan tanam berupa bibit, untuk itu perlu tindakan kultur teknis atau perawatan bibit yang baik [3]. Pemilihan bibit yang tepat akan mempengaruhi produktifitas tanaman kelapa sawit. Menurut Septianita faktor bibit berpengaruh sangat nyata terhadap produksi kelapa sawit dengan koefisien regresi produksi sebesar 2,267 berarti penambahan satu persen bibit akan meningkatkan produksi sebesar 226,7 persen, jika ditinjau dari elastisitasnya lebih dari 1 ($0 < EP < 1$) yang berarti penambahan bibit meningkatkan produksi 226,7 persen [4]. Dalam penentuan jenis bibit permasalahan yang sering dihadapi adalah kesesuaian jenis bibit terhadap kondisi daerah tanam dan jenis perawatan. Daerah tanam yang digunakan untuk menanam kelapa sawit memiliki karakteristik yang beragam. Selain itu, perawatan pada tanaman kelapa sawit juga berbeda-beda pada tiap daerah tanam. Kebanyakan pekebun kelapa sawit memilih jenis bibit kelapa sawit tanpa mempertimbangkan dengan kesesuaian daerah tanam serta perawatannya. Hal ini menyebabkan produktifitas kelapa sawit yang ditanam tidak maksimal.

Melihat perusahaan-perusahaan perkebunan kelapa sawit di Indonesia khususnya didaerah Kabupaten Landak, Kalimantan Barat yang berkembang sangat pesat tentunya sangat membantu dalam menganalisa jenis bibit kelapa sawit apa yang tepat pada kondisi lahan dan jenis perawatan tertentu. Hal ini dikarenakan Perusahaan-perusahaan perkebunan kelapa sawit tersebut memiliki lahan yang sangat luas dengan data-data kondisi daerah tanam yang berbeda-beda serta perawatan tanaman yang beragam. Data-data tersebut dapat dijadikan referensi dalam penentuan jenis bibit yang tepat. Salah satu teknologi yang dapat memanfaatkan data-data tersebut adalah Machine learning. Machine learning adalah teknologi yang mampu mempelajari data yang ada dan

melakukan tugas-tugas tertentu sesuai dengan apa yang telah pelajari.

Machine learning merupakan sebuah studi tentang algoritma untuk mempelajari sesuatu dalam melakukan beberapa hal tertentu yang dilakukan oleh manusia secara otomatis. Belajar dalam hal ini berkaitan dengan bagaimana menuntaskan berbagai tugas yang ada, atau membuat suatu prediksi kesimpulan baru yang akurat dari berbagai pola yang sudah dipelajari sebelumnya. Algoritma Machine learning terdiri dari beberapa jenis, pada kasus ini algoritma yang digunakan adalah naive bayes. Berdasarkan penelitian terkait oleh Welmy Sinlae, Sebastianus A. S. Mola dan Nelci D. Rumlaklak mengenai penentuan kesesuaian lahan pertanian tanaman cabai menggunakan metode naive bayes di kabupaten kupang. Dalam penelitian ini penggunaan 7 kriteria dan sub kriterianya dalam penentuan kesesuaian lahan sudah tepat. Hasil pengujian menunjukkan hasil yang memuaskan dengan rata-rata akurasi sebesar 96,48%. Demikian juga jika urutan waktu pengambilan data diabaikan, dengan pengujian akurasi menggunakan 10-fold cross validation menunjukkan hasil juga memuaskan sebesar 89,1%. Adapun penelitian lain yang dilakukan oleh Galih Dapa Imanda, Ramadiani dan Indah Fitri Astuti mengenai Penerapan Metode Naive Bayes Untuk Kesesuaian Lahan Tanaman Jagung. Pada penelitian ini Metode Naive Bayes berhasil diterapkan dalam sistem penunjang keputusan Penentuan Kesesuaian Lahan Tanaman Jagung dengan kriteria Temperatur/suhu, curah hujan, jenis tanah, dan sinar matahari dapat diproses dan kemudian menghasilkan rekomendasi melalui tahap-tahap yang telah ditetapkan dalam metode Naive Bayes.

Berdasarkan uraian diatas maka penelitian ini akan menggunakan metode naive bayes untuk melakukan Penentuan Bibit Kelapa Sawit Berdasarkan Kondisi Daerah Tanam dan Perawatan Tanaman. Studi Kasus dilakukan di PT. Satria Multi Sukses, Kab. Landak. Dengan ini diharapkan dapat mempercepat dan mempermudah dalam memilih bibit kelapa sawit yang sesuai dengan kriteria daerah yang akan ditanam dalam upaya meningkatkan produktifitas kelapa sawit. Naive bayes merupakan metode pengklasifikasian ada tidaknya ciri tertentu dari sebuah kelas. Penentuan bibit tanaman kelapa sawit dilihat dari hasil panen yang menunjukkan kesesuaian bibit terhadap kondisi lahan tanam serta jenis perawatan yang diberikan terhadap tanaman. Penentuan jenis bibit tanaman kelapa sawit dengan metode ini membutuhkan informasi probabilitas setiap alternatif pada persoalan yang dihadapi untuk menghasilkan nilai harapan sebagai bibit yang baik.

Penelitian terkait dengan penelitian ini, diantaranya oleh: (a) Sebastianus A. S. Mola dan Nelci D. Rumlaklak mengenai penentuan kesesuaian lahan pertanian tanaman cabai menggunakan metode naive bayes di Kabupaten Kupang. Dalam penelitian ini penggunaan 7 kriteria yaitu curah hujan, drainase, tekstur tanah, kedalaman tanah, C-organik, kemiringan lahan dan bahaya terjadinya bencana. Sistem ini dapat membantu dalam menentukan kesesuaian

lahan pertanian untuk penanaman cabai [5]. (b) Galih Dapa Imanda, Ramadiani dan Indah Fitri Astuti membahas tentang Penerapan Metode Naive Bayes Untuk Kesesuaian Lahan Tanaman Jagung. Pada penelitian ini Metode Naive Bayes berhasil diterapkan dalam sistem penunjang keputusan Penentuan Kesesuaian Lahan Tanaman Jagung dengan kriteria Temperatur/suhu, curah hujan, jenis tanah, dan sinar matahari dapat diproses dan kemudian menghasilkan rekomendasi melalui tahap-tahap yang telah ditetapkan dalam metode Naive Bayes [6]. (c) Sri Rahayu dan Anita Sindar RM Sinaga membahas Pemilihan Kualitas Jenis Rumput Taman CV. Rumput Kita Lanscape menggunakan metode klasifikasi naive bayes. Hasilnya sistem dapat mempertimbangkan konsistensi yang logis dalam penilaian yang digunakan sehingga menghasilkan alternatif yang tidak banyak dan mampu memberikan keputusan tentang pemilihan kualitas jenis rumput yang tepat [7]. (d) Alfa Saleh Menentukan konsentrasi siswa di MAS PAB 2 Medan menggunakan Metode Naive Bayes. Metode Naive Bayes memanfaatkan data training untuk menghasilkan probabilitas setiap kriteria untuk class yang berbeda, sehingga nilai-nilai probabilitas dari kriteria tersebut dapat dioptimalkan untuk memprediksi konsentrasi siswa berdasarkan proses klasifikasi yang dilakukan oleh metode Naive Bayes itu sendiri [8]. Selain itu, Sharly dan Asrul Ashari Muin menggunakan metode naive bayes dalam memprediksi kelulusan pada perguruan tinggi. Pada penelitian ini menunjukkan metode naive bayes dalam menentukan potensi masuk pada salah satu pilihan lulus dengan tingkat akurasi 94% [9]. (e) Sri Widaningsih membandingkan metode data mining untuk prediksi nilai dan waktu kelulusan mahasiswa prodi teknik informatika dengan algoritma c4.5, naive bayes, knn, dan svm. Dari hasil evaluasi diperoleh hasil bahwa algoritma naive bayes yang paling baik untuk memprediksi tingkat kelulusan yang diharapkan karena memiliki nilai akurasi tertinggi dan error terkecil dibandingkan dengan algoritma lainnya [10]. (f) Xhemali, Daniela, Chris J. Hinde, and Roger G. Stone yang berjudul "Naive Bayes vs. decision trees vs. neural networks in the classification of training web pages.", mengatakan bahwa "Naive Bayes Classifier memiliki tingkat akurasi yg lebih baik dibanding model classifier lainnya" [11].

A. Kelapa Sawit

Kelapa sawit merupakan tanaman dengan nilai ekonomis yang cukup tinggi karena merupakan salah satu tanaman penghasil minyak nabati. Bagi Indonesia, kelapa sawit memiliki arti penting karena mampu menciptakan kesempatan kerja bagi masyarakat dan sebagai sumber perolehan devisa negara. Sampai saat ini Indonesia merupakan salah satu produsen utama minyak sawit dunia selain Malaysia dan Nigeria [12] Kelapa sawit merupakan tanaman monokotil, yaitu batangnya tidak mempunyai kambium dan umumnya tidak bercabang. Batang berfungsi sebagai penyangga tajuk serta menyimpan dan mengangkat bahan makanan.

B. Machine learning

Machine learning merupakan sebuah studi tentang algoritma untuk mempelajari sesuatu dalam melakukan beberapa hal tertentu yang dilakukan oleh manusia secara otomatis. Belajar dalam hal ini berkaitan dengan bagaimana menuntaskan berbagai tugas yang ada, atau membuat suatu prediksi kesimpulan baru yang akurat dari berbagai pola yang sudah dipelajari sebelumnya [13]. *Machine learning* terbagi menjadi 2 macam konsep pembelajaran, yaitu pertama Supervised Learning yang merupakan teknik pembelajaran mesin yang membuat suatu fungsi berdasarkan data latihan yang sudah ada, dalam hal ini dapat dikatakan untuk teknik ini sudah tersedia data latihan secara detil dan terklasifikasi dengan baik yang akan dijadikan sebuah model data saat dilakukan proses ujicoba dengan data tes yang baru dan menghasilkan hasil keluaran yang sesuai diharapkan sebelumnya berdasarkan data latihan yang ada. Kedua adalah Unsupervised Learning yang merupakan teknik pembelajaran mesin yang berusaha untuk melakukan representasi pola sebuah input yang berasal dari data latihan dan salah satu yang menjadi perbedaan dengan Supervised Learning adalah tidak adanya pengklasifikasian dari input data.

C. Klasifikasi

Klasifikasi merupakan salah satu dari Supervised Learning. Klasifikasi merupakan proses untuk menemukan fungsi dan model yang dapat membedakan atau menjelaskan konsep atau kelas data dengan tujuan memperkirakan kelas yang tidak diketahui dari suatu objek. Dalam proses pengklasifikasian biasa terdapat dua proses yang harus dilakukan, yaitu [14] :

1. Proses Training

Pada proses ini akan digunakan data training set atau data sampel yang telah diketahui label – label atau atribut dari data sampel tersebut untuk membangun model.

2. Proses Testing

Pada proses testing ini dilakukan untuk mengetahui keakuratan model yang telah dibuat pada proses training maka dibangun data yang disebut dengan data testing untuk mengklasifikasi label – labelnya.

D. Metode Naive Bayes

Naive Bayes merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman dimasa sebelumnya. Naive Bayes didasarkan pada asumsi penyederhanaan bahwa nilai atribut secara kondisional saling bebas jika diberikan nilai output. Dengan kata lain, diberikan nilai output, probabilitas mengamati secara bersama adalah produk dari probabilitas individu. Algoritma ini mengasumsikan bahwa atribut obyek adalah independen. Probabilitas yang terlibat dalam memproduksi perkiraan akhir dihitung sebagai jumlah frekuensi dari "master" tabel keputusan.

Pengklasifikasi Bayesian memiliki tingkat kesalahan minimal dibandingkan dengan klasifikasi lainnya. Namun,

dalam prakteknya hal ini tidak selalu terjadi, karena ketidakakuratan asumsi yang dibuat untuk penggunaannya, seperti kondisi kelas independen, dan kurangnya data probabilitas yang tersedia. Pengklasifikasi Bayesian juga berguna dalam memberikan pembenaran teoritis untuk pengklasifikasi lain yang tidak secara eksplisit menggunakan teorema Bayes.

E. SMOTE

SMOTE merupakan teknik menyeimbangkan jumlah distribusi data sampel pada kelas minoritas dengan cara menyeleksi data sampel tersebut hingga jumlah data sampel menjadi seimbang dengan jumlah sampel pada kelas mayoritas [15]. Konsep SMOTE yaitu membuat replica dari class minoritas sehingga mirip dengan class mayoritas, replica ini disebut sintetis [16].

II. METODOLOGI PENELITIAN

Cara mudah agar format makalah Anda sesuai dengan format makalah yang kami perlukan, gunakan dokumen ini sebagai *template* dan ketik teks Anda di dalamnya.

Beberapa tahapan dalam penelitian (lihat Gambar 1) yang memenuhi tugas akhir ini adalah sebagai berikut:

1) Pengumpulan Data

Pengumpulan data dilakukan untuk mendapatkan informasi dan kebutuhan data untuk proses pemodelan. Pengumpulan data dilakukan melalui wawancara dan observasi terhadap perkebunan kelapa sawit PT. Satria Multi Sukses, Kab. Landak.

2) Pengelolaan Data Set

Dataset yang didapatkan akan mengalami proses pembangkitan data pemilihan data yang akan digunakan untuk proses selanjutnya. Pengelolaan dataset terdiri dari seleksi data, penentuan atribut, pelabelan data dan inialisasi data.

3) Preprocessing Data

Preprocessing data yang dilakukan pada penelitian ini meliputi:

- Cleaning

Pada penelitian ini, proses Cleaning yang dilakukan adalah menghapus Duplikat data, menghapus kolom yang tidak representatif dan menghapus kolom yang hanya mempunyai nilai tunggal.

- Ordinal encoding

Ordinal encoding adalah mengubah fitur bertipe kategori menjadi bilangan dengan Ordinal encoding.

4) Pemodelan

Pendekatan yang dipilih untuk melakukan klasifikasi pada penelitian ini yaitu Metode Naive Bayes. Model dibangun menggunakan IDE Jupyter Notebook dengan menggunakan bahasa pemrograman python.

5) Evaluasi

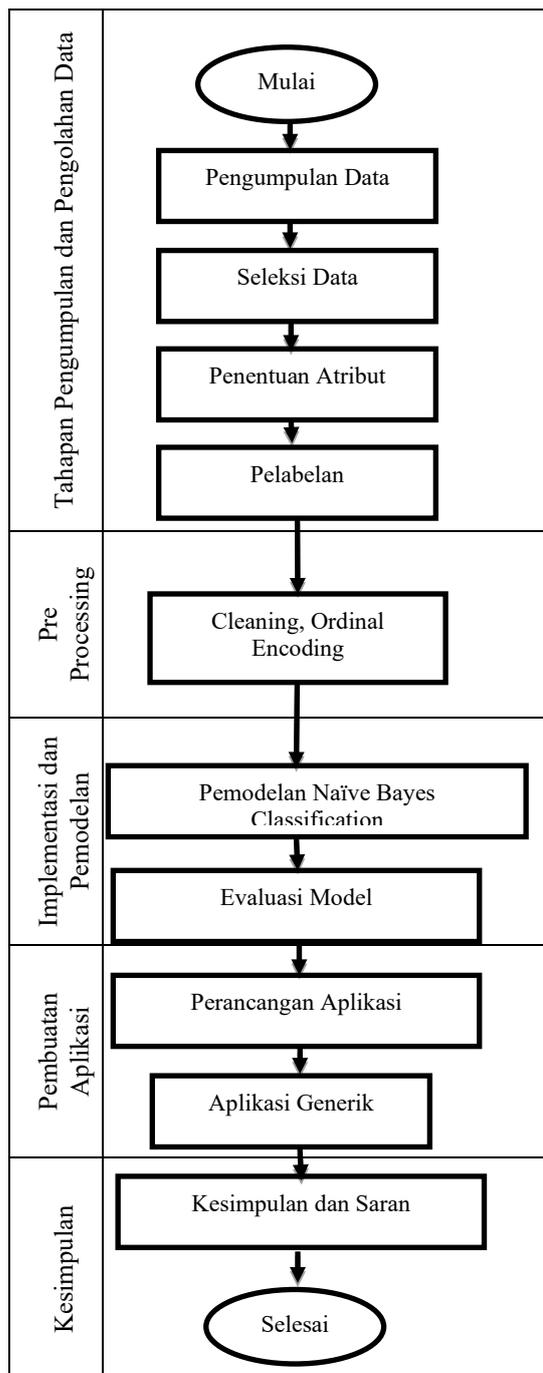
Evaluasi dilakukan untuk melihat hasil akurasi, presisi dan sensitivitas klasifikasi dari pengujian data uji.

6) Pembuatan aplikasi

Tahapan selanjutnya adalah membuat aplikasi dengan model yang telah dibuat sebelumnya. Platform yang digunakan dalam pembuatan aplikasi ini adalah Visual Studio Code dengan bahasa pemrograman python.

7) Kesimpulan

Penarikan kesimpulan didapatkan setelah melakukan analisis hasil dari pengujian atau evaluasi, sehingga dapat diambil kesimpulan mengenai model yang dibuat apakah menunjukkan keakuratan yang tinggi atau tidak.



Gambar 1. Metodologi penelitian

III. HASIL DAN PEMBAHASAN

Tahap-tahap yang dilakukan untuk membuat model *Machine learning* mulai dari pengumpulan data dan pengolahan data, preprocessing, pemodelan hingga pembuatan aplikasi. Terdapat empat model dibuat pada penelitian ini. Data uji pada setiap model dibagi menggunakan 10-Fold Cross Validation. Hasil Prediksi data uji dibandingkan dengan label untuk dievaluasi menggunakan Confusion Matrix.

A. Pengumpulan dan Pengolahan Data

1) Pengumpulan Data

Pada penelitian ini, terdapat dua jenis data yang digunakan, diantaranya adalah data primer dan data sekunder. Data primer adalah data yang pertama kali diperoleh secara langsung dari objek penelitian. Pada penelitian ini, yang menjadi data primer adalah Daftar Blok Lokasi Perkebunan Kelapa Sawit, Daftar Kondisi Lahan, Daftar Perawatan dan Daftar Hasil Panen Perkebunan kelapa sawit PT. Satria Multi Sukses, Kab. Landak. Data sekunder adalah data yang tidak diperoleh langsung dari objek penelitian, yaitu merupakan hasil studi pustaka dan referensi mengenai teori-teori dan ilmu pengetahuan yang mendukung sebuah penelitian. Referensi yang digunakan pada penelitian ini adalah buku, jurnal, serta referensi dari internet.

2) Pengolahan Data

Dalam menentukan atribut untuk proses klasifikasi penentuan bibit tanaman kelapa sawit mengacu dengan Pedoman Budidaya Kelapa Sawit (*Elais Guineensis*) Yang Baik Oleh Direktorat Jenderal Perkebunan Kementerian Pertanian 2014 dan Panduan Teknis Budidaya Kelapa Sawit Untuk Praktisi Perkebunan Oleh Iyung Pahan. Informasi umum yang dijadikan atribut ialah kondisi lahan, perawatan dan hasil panen pada perkebunan sawit PT. Satria Multi Sukses, Kab. Landak. Atribut tersebut diantaranya Jenis Tanah, PH Tanah, Jenis Dataran, Perawatan Pelepah (Pruning), Pengendalian Gulma, Perawatan Piringan dan Pemupukan.

Pada Gambar 2 berikut ini merupakan dokumen atau data dari tiap blok beserta atributnya dan disimpan dalam bentuk file .xlsx. Data-data ini selanjutnya akan dilanjutkan ke tahap pelabelan data.

NamaBlok	JenisTanah	JenisDataran	KondisiDataran	PHTanah	KelasDrainase	Pruning	Pemupukan	Penyakit	BTP	Babun	Urea	MOP	Zinc	Copp	Rock	HGF	TSP			
21110A120	Pasir	Rawa-rawa	Datar	9	Lancar	1,00	3,00	2,00	4,00	2,00	2,00	1,00	1,50	1,00	0,05	0,05	0,75	0,00	0,00	0,00
1S200C120	Pasir	DataranRendah	Datar	7	TidakAda	1,00	3,00	2,00	4,00	1,00	2,00	1,00	1,50	0,00	0,05	0,05	0,75	0,00	0,00	0,00
2S200C120	Pasir	DataranRendah	Datar	7	TidakAda	1,00	3,00	2,00	4,00	1,00	2,00	1,00	1,50	0,00	0,05	0,05	0,75	0,00	0,00	0,00
3S200C120	Pasir	DataranRendah	Datar	7	TidakAda	1,00	3,00	2,00	4,00	1,00	2,00	1,00	1,50	0,00	0,05	0,05	0,75	0,00	0,00	0,00
4S200C120	Pasir	DataranRendah	Datar	7	TidakAda	1,00	3,00	2,00	4,00	1,00	2,00	1,00	1,50	0,00	0,05	0,05	0,75	0,00	0,00	0,00
5S200C120	Pasir	DataranRendah	Datar	7	TidakAda	1,00	3,00	2,00	4,00	1,00	2,00	1,00	1,50	0,00	0,05	0,05	0,75	0,00	0,00	0,00
2112A120	Pasir	Rawa-rawa	Datar	9	Lancar	1,00	2,00	2,00	4,00	1,00	2,00	1,00	1,50	0,00	0,05	0,05	0,75	0,00	0,00	0,00
3112A120	Pasir	Rawa-rawa	Datar	9	Sedang	1,00	2,00	2,00	4,00	1,00	2,00	1,00	1,50	0,00	0,05	0,05	0,75	0,00	0,00	0,00
4112A120	Pasir	Rawa-rawa	Datar	9	Sedang	1,00	2,00	2,00	4,00	1,00	2,00	1,00	1,50	0,00	0,05	0,05	0,75	0,00	0,00	0,00
1112A120	Pasir	Rawa-rawa	Datar	9	Sedang	1,00	2,00	2,00	4,00	1,00	2,00	1,00	1,50	0,00	0,05	0,05	0,75	0,00	0,00	0,00
2112A120	Pasir	Rawa-rawa	Datar	9	Lancar	1,00	2,00	2,00	4,00	1,00	2,00	1,00	1,50	0,00	0,05	0,05	0,75	0,00	0,00	0,00
3112A120	Pasir	Rawa-rawa	Datar	9	Lancar	1,00	2,00	2,00	4,00	1,00	2,00	1,00	1,50	0,00	0,05	0,05	0,75	0,00	0,00	0,00
1S1A120	Mineral	DataranRendah	Miring	5	TidakAda	2,00	3,00	1,00	4,00	2,00	2,00	1,00	1,50	0,00	0,00	0,00	0,75	0,00	0,00	0,00
1S2A120	Mineral	DataranRendah	Miring	8	TidakAda	2,00	2,00	1,00	4,00	1,00	2,00	1,00	1,50	0,00	0,00	0,00	0,75	0,00	0,00	0,00
2S2A120	Mineral	DataranRendah	Miring	9	TidakAda	2,00	2,00	1,00	4,00	1,00	2,00	1,00	1,50	0,00	0,00	0,00	0,75	0,00	0,00	0,00

Gambar 2. Data Atribut

Data hasil pelabelan akan dimuat kedalam sebuah tabel (lihat contoh Tabel 1) dan disimpan kedalam bentuk

data .CSV, untuk dijadikan sebagai data set dalam klasifikasi.

TABEL I
PELABELAN

Label	Nama Blok
SRJ	1R25A120
ASD	S24A120
ASD	1S11A120
SRJ	10R25A120

Pada Gambar 3 berikut ini merupakan dokumen atau data set yang telah dilabel dalam bentuk file .csv.

Gambar 3. Data Set

Dataset merupakan hasil dari tahapan seleksi data dan pelabelan data dimana data tersebut adalah data yang telah disusun rapi dan relevan dalam penelitian.

B. Hasil Implementasi Preprocessing

Pada penelitian ini tahapan Preprocessing terbagi menjadi dua diantaranya adalah tahapan Cleaning dan tahapan Encoding. Data awal yang telah melalui tahapan preprocessing kemudian disimpan (lihat Kode Program 3.2) menjadi data baru dalam format .CSV untuk kemudian dipakai dalam proses pemodelan.

1) Cleaning

Terdapat 3 proses cleaning yang dilakukan diantaranya menghapus Duplikat data, menghapus kolom yang tidak representatif dan menghapus kolom yang hanya mempunyai nilai tunggal pada penelitian ini.

2) Encoding

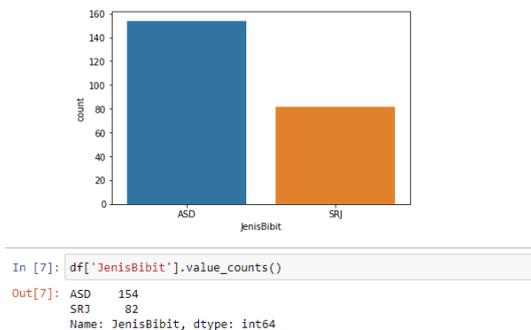
Data-data yang diubah menggunakan Ordinal encoding diantaranya ialah data jenis tanah, data jenis dataran, data kondisi dataran dan data kelas drainase. Output setelah melakukan Ordinal Encoding dapat dilihat pada Gambar 4.

Out[10]:	JenisBlok	JenisTanah	JenisDataran	KondisiDataran	PHTanah	KelasDrainase	Pruning	Pengendalianalang	GarakPiringan	BTP	Urea(kg)	MOP(kg)	0	1	2	3	4	5
0	ASD	1	1	1	10	1	10	20	10	10	10	15						
1	ASD	1	1	2	10	1	10	20	10	10	10	15						
3	ASD	2	1	1	10	1	10	20	10	10	10	15						
4	ASD	1	1	1	8	1	10	20	10	20	10	15						
5	ASD	1	1	2	8	1	10	20	10	20	10	15						
802	ASD	1	2	2	10	2	10	30	20	20	10	27						
809	ASD	2	1	1	11	1	20	20	10	20	10	15						
810	ASD	2	1	2	8	2	10	20	10	20	10	15						
811	ASD	1	1	6	1	10	10	20	10	20	10	15						

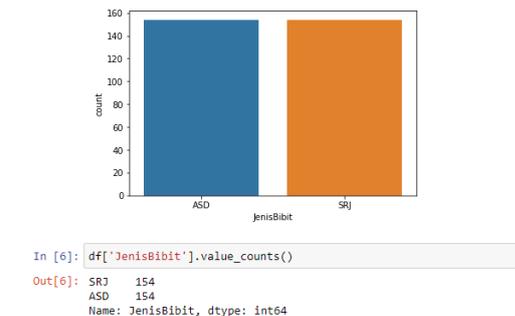
Gambar 4. Hasil Ordinal Encoding

C. Hasil Penerapan SMOTE

Total keseluruhan dataset sebelum dilakukan SMOTE adalah 236 data (Datasetpre.CSV) yang terdiri dari 154 pada kelas ASD dan 82 pada kelas SRJ (lihat Gambar 5). Setelah dilakukan tahapan SMOTE, kelas ASD dan SRJ menjadi seimbang dengan total keseluruhan data menjadi 308 data yang terdiri dari 154 data pada kelas ASD dan 154 pada kelas SRJ (lihat Gambar 6).



Gambar 5. Dataset Awal



Gambar 6. Dataset dengan SMOTE

Data hasil SMOTE disimpan menjadi dataset baru dalam bentuk file .CSV yang akan digunakan pada pemodelan.

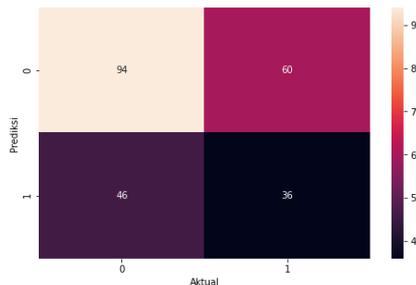
D. Hasil Evaluasi Gaussian Naive Bayes

Pada penelitian ini terdapat dua skenario pengujian terhadap dua jenis model naïve bayes yaitu Gaussian Naïve Bayes dan Multinomial Naïve Bayes. Skenario pengujian pertama dilakukan dengan menggunakan algoritma default klasifikasi dan skenario pengujian kedua dilakukan dengan menggunakan algoritma default aplikasi dengan penambahan teknik Data Balancing SMOTE (Syntetic Minority Over Sampling Technique) dalam penanganan Imbalance Data.

TABEL II
SKENARIO PENGUJIAN

Skenario Pengujian	SMOTE
Skenario Pengujian Pertama	X
Skenario Pengujian Kedua	√

Confusion matrix untuk model pertama gaussian naive bayes dapat dilihat pada Gambar 7.



Gambar 7. Confusion Matrix Model Pertama Gaussian Naive Bayes

Pengolahan ini menggunakan data sebanyak 236 record. Parameter pengujian yang dilihat yaitu Accuracy, Precision, Recall dan f1-score.

1) Accuracy

$$\begin{aligned} \text{Accuracy} &= (TP+TN) / (TP+TN+FP+FN) \\ &= (94+36) / (94+36+60+46) \\ &= 0,55 \end{aligned}$$

2) Precision

$$\begin{aligned} \text{Precision} &= TP / (TP+FP) \\ &= 94 / (94+60) \\ &= 0,61 \end{aligned}$$

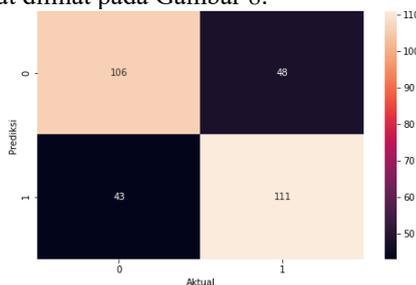
3) Recall

$$\begin{aligned} \text{Recall} &= TP / (TP+FN) \\ &= 94 / (94+46) \\ &= 0,67 \end{aligned}$$

4) F-1 Score

$$\begin{aligned} \text{F-1 Score} &= (2 \times \text{Recall} \times \text{Precision}) / (\text{Recall} \\ &\quad + \text{Precision}) \\ &= (2 \times 0,67 \times 0,61) / (0,67 + 0,61) \\ &= 0,82 / (1,28) \\ &= 0,63 \end{aligned}$$

Confusion matrix untuk model kedua gaussian naive bayes dapat dilihat pada Gambar 8.



Gambar 8. Confusion Matrix Model Kedua Gaussian Naive Bayes

Pengolahan ini menggunakan data sebanyak 308 record. Parameter pengujian yang dilihat yaitu Accuracy, Precision, Recall dan f1-score.

1) Accuracy

$$\begin{aligned} \text{Accuracy} &= (TP+TN) / (TP+TN+FP+FN) \\ &= (106+111) / (106+111+48+43) \\ &= 0,70 \end{aligned}$$

2) Precision

$$\begin{aligned} \text{Precision} &= \text{TP} / (\text{TP}+\text{FP}) \\ &= 106 / (106+48) \\ &= 0,68 \end{aligned}$$

3) Recall

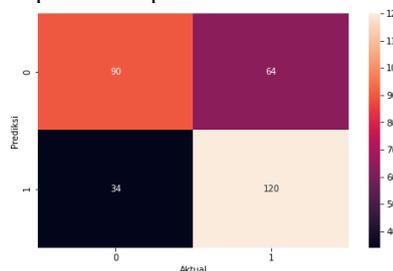
$$\begin{aligned} \text{Recall} &= \text{TP} / (\text{TP}+\text{FN}) \\ &= 106 / (106+43) \\ &= 0,71 \end{aligned}$$

4) F-1 Score

$$\begin{aligned} \text{F-1 Score} &= (2 \times \text{Recall} \times \text{Precision}) / (\text{Recall} \\ &+ \text{Precision}) \\ &= (2 \times 0,71 \times 0,68) / (0,71 + 0,68) \\ &= 0,97 / (1,39) \\ &= 0,69 \end{aligned}$$

$$\begin{aligned} &= (2 \times 0,75 \times 0,61) / (0,75 + 0,61) \\ &= 0,91 / (1,36) \\ &= 0,66 \end{aligned}$$

Confusion matrix untuk model kedua multinomial naive bayes dapat dilihat pada Gambar 10.



Gambar 10. Confusion Matrix Model Kedua Multinomial Naive Bayes

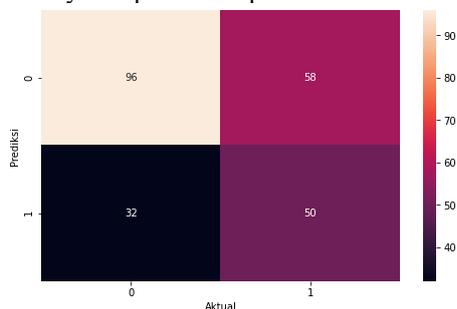
Berikut merupakan hasil evaluasi dari kedua model Gaussian Naive Bayes dijabarkan kedalam Tabel 3.

TABEL III
HASIL EVALUASI MODEL GAUSSIAN NAIVE BAYES

Parameter Pengujian	Model Pertama	Model Kedua
Accuracy	55 %	70 %
Precision	61 %	68 %
Recall	67 %	71 %
F1-Score	63 %	69 %

E. Hasil Evaluasi Model Multinomial Naive Bayes

Confusion matrix untuk model pertama Multinomial Naive Bayes dapat dilihat pada Gambar 9.



Gambar 9. Confusion Matrix Model Pertama Multinomial Naive Bayes

Pengolahan ini menggunakan data sebanyak 236 record. Parameter pengujian yang dilihat yaitu Accuracy, Precision, Recall dan f1-score.

1. Accuracy

$$\begin{aligned} \text{Accuracy} &= (\text{TP}+\text{TN}) / (\text{TP}+\text{TN}+\text{FP}+\text{FN}) \\ &= (96+50) / (96+50+58+32) \\ &= 0,62 \end{aligned}$$

2. Precision

$$\begin{aligned} \text{Precision} &= \text{TP} / (\text{TP}+\text{FP}) \\ &= 96 / (96+58) \\ &= 0,61 \end{aligned}$$

3. Recall

$$\begin{aligned} \text{Recall} &= \text{TP} / (\text{TP}+\text{FN}) \\ &= 96 / (96+32) \\ &= 0,75 \end{aligned}$$

4. F-1 Score

$$\begin{aligned} \text{F-1 Score} &= (2 \times \text{Recall} \times \text{Precision}) / (\text{Recall} \\ &+ \text{Precision}) \end{aligned}$$

Pengolahan ini menggunakan data sebanyak 308 record. Parameter pengujian yang dilihat yaitu Accuracy, Precision, Recall dan f1-score.

1) Accuracy

$$\begin{aligned} \text{Accuracy} &= (\text{TP}+\text{TN}) / (\text{TP}+\text{TN}+\text{FP}+\text{FN}) \\ &= (90+120) / (90+120+64+34) \\ &= 0,68 \end{aligned}$$

2) Precision

$$\begin{aligned} \text{Precision} &= \text{TP} / (\text{TP}+\text{FP}) \\ &= 90 / (90+64) \\ &= 0,58 \end{aligned}$$

3) Recall

$$\begin{aligned} \text{Recall} &= \text{TP} / (\text{TP}+\text{FN}) \\ &= 90 / (90+34) \\ &= 0,72 \end{aligned}$$

4) F-1 Score

$$\begin{aligned} \text{F-1 Score} &= (2 \times \text{Recall} \times \text{Precision}) / (\text{Recall} \\ &+ \text{Precision}) \\ &= (2 \times 0,72 \times 0,58) / (0,72 + 0,58) \\ &= 0,84 / (1,30) \\ &= 0,64 \end{aligned}$$

Berikut merupakan hasil evaluasi dari kedua Model Multinomial Naive Bayes dijabarkan kedalam Tabel 4.

TABEL IV
HASIL EVALUASI MODEL MULTINOMIAL NAIVE BAYES

Parameter Pengujian	Model Pertama	Model Kedua
Accuracy	62 %	68 %
Precision	61 %	58 %
Recall	75 %	72 %
F1-Score	66 %	64 %

F. Analisis Gaussian Naive Bayes

Hasil evaluasi model gaussian naive bayes (Lihat Tabel 4.1) menunjukkan hasil dari model pertama gaussian naive bayes memperoleh hasil performa dengan nilai Accuracy, Precision, Recall dan F1-Score masing-masing 55 %, 61%, 67% dan 63 %. Confusion Matrix model pertama gaussian naive bayes dapat dilihat pada

Gambar 4.1. Model pertama gaussian naive bayes menggunakan dataset awal dimana dataset tersebut mempunyai data kelas yang tidak seimbang. Dari 236 record data, 154 record berada pada kelas ASD dan 82 record berada pada kelas SRJ. Hasil evaluasi model pertama menunjukkan bahwa performa dari gaussian naive bayes lemah terhadap dataset yang mempunyai kelas data yang tidak seimbang.

Sedangkan pada model kedua memperoleh hasil dengan nilai Accuracy Precision, Recall dan F1-Score masing-masing 70%, 68%, 71% dan 69%. Hasil evaluasi model kedua gaussian naive bayes menunjukkan bahwa penambahan teknik data balancing SMOTE memberikan pengaruh kenaikan Accuracy, Precision, Recall dan F1-Score yaitu masing-masing mengalami kenaikan sebesar 15%, 7%, 3% dan 6%. Confusion Matrix model kedua gaussian naive bayes dapat dilihat pada Gambar 4.2. Pada Confusion Matrix Model Kedua gaussian naive bayes menunjukkan bahwa pada kelas SRJ terjadi kenaikan jenis bibit yang terprediksi dengan benar sehingga mengakibatkan nilai akurasi meningkat.

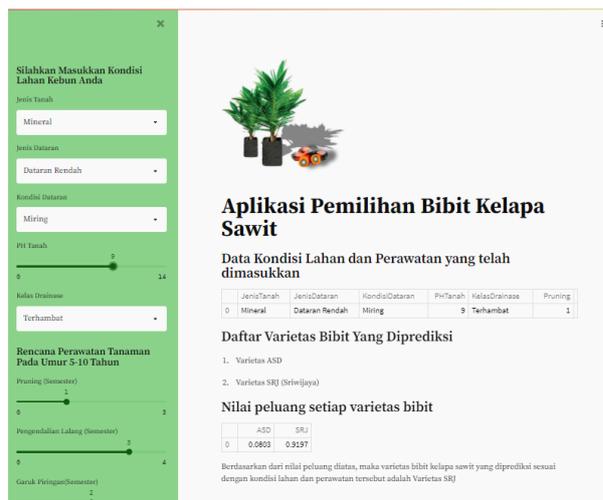
G. Analisis Multinomial Naive Bayes

Hasil evaluasi model gaussian naive bayes (Lihat Tabel 4.2) menunjukkan hasil dari model pertama multinomial naive bayes memperoleh hasil performa dengan nilai Accuracy, Precision, Recall dan F1-Score masing-masing 62 %, 61%, 75% dan 66 %. Confusion Matrix model pertama multinomial naive bayes dapat dilihat pada Gambar 4.3. Confusion Matrix model pertama multinomial naive bayes menunjukkan kesalahan paling banyak terjadi pada kelas SRJ, yang terprediksi oleh mesin sebagai kelas ASD.

Sedangkan pada model kedua memperoleh hasil dengan nilai Accuracy Precision, Recall dan F1-Score masing-masing 68%, 58%, 72% dan 64%. Hasil evaluasi model kedua multinomial naive bayes menunjukkan bahwa penambahan teknik data balancing SMOTE memberikan pengaruh kenaikan terhadap Accuracy yaitu mengalami kenaikan sebesar 6%, sedangkan nilai Precision, Recall dan F1-Score mengalami penurunan masing-masing sebesar 3 %, 3% dan 2%. Confusion Matrix model kedua multinomial naive bayes dapat dilihat pada Gambar 4.4. Hasil evaluasi model kedua multinomial naive bayes menunjukkan bahwa penerapan teknik balancing data SMOTE pada multinomial naive bayes tidak berpengaruh positif terhadap nilai Precision, Recall dan F1-Score..

H. Implementasi aplikasi generik

Aplikasi generik dibangun berdasarkan model dengan hasil evaluasi yang paling baik yaitu Gaussian Naive Bayes dan data balancing SMOTE dengan nilai akurasi sebesar 70%. Aplikasi generik yang dibangun berbasis web dengan fitur kolom input kondisi lahan dan jenis perawatan yang dapat dipilih maupun diisi dengan angka. Kemudian setelah kondisi lahan dan jenis perawatan diinput, aplikasi akan menampilkan output berupa bibit yang sesuai. Tampilan aplikasi generik dapat dilihat pada Gambar 11.



Gambar 11. Tampilan Aplikasi

IV. KESIMPULAN

Berdasarkan hasil penelitian yang telah diuraikan, maka dapat ditarik kesimpulan.

Terdapat empat buah model naive bayes yang dapat diterapkan untuk penentuan bibit kelapa sawit berdasarkan kondisi daerah tanam dan perawatan tanaman diantaranya adalah dua buah model gaussian naive bayes dan dua buah model multinomial naive bayes.

Adanya bentuk dan fungsi aplikasi yang dibangun berdasarkan model dengan hasil evaluasi yang paling baik yaitu Gaussian Naive Bayes model kedua yang dapat dipakai untuk masyarakat dalam penentuan bibit kelapa sawit sehingga dapat membantu dalam memprediksi bibit yang sesuai dengan kondisi lahan dan perawatan tanaman.

DAFTAR PUSTAKA

- [1] Pahan, I. 2012. "Panduan Lengkap Kelapa Sawit", Penebar Swadya. Jakarta.
- [2] Khaswarina, S. 2001. "Keragaan bibit kelapa sawit terhadap pemberian berbagai kombinasi pupuk di pembibitan utama". Jurnal Natur Indonesia Volume 3 No.2.
- [3] Haekal, M. 2000. "Respon pertumbuhan bibit kelapa sawit terhadap pemupukan pada media tumbuh alang-alang dengan inokulasi *Trichoderma viride*" <http://repository.ipb.ac.id>. Diakses tanggal 03 Agustus 2020.
- [4] Septianita 2009. Faktor-Faktor yang Mempengaruhi Produksi Kelapa Sawit (*Elaeis quinensis* Jack) dan Kontribusinya Terhadap Pendapatan Keluarga di Desa Makartitama Kec. Peninjauan Kab. OKU. Jurnal Agroteknologi. 1(2): 78 – 85.
- [5] Sinlae, Welmy, A. S Sebastianus. Mola dan Rumlaklak D. Nelci. 2021. "Penentuan Kesesuaian Lahan Pertanian Tanaman Cabai Menggunakan Metode Naive Bayes Di Kabupaten Kupang". J-ICON, Vol. 9 No. 1 pp. 56~64
- [6] Imanda, Galih Dapa, Ramadiani dan Astuti, Indah Fitri, 2020. "Penerapan Metode Naive Bayes Untuk Kesesuaian Lahan Tanaman Jagung". Prosiding Seminar Nasional Ilmu Komputer dan Teknologi Informasi Vol. 5, No. 1. e-ISSN 2540-7902
- [7] Rahayu, Sri dan Sinaga, Anita Sinder. 2018. "Penerapan Metode Naive Bayes Dalam Pemilihan Kualitas Jenis Rumput Taman CV. Rumput Kita Landscape". Jurnal Teknologi Informasi & Komunikasi Digital Zone 9 (2), 162-173,
- [8] SALEH, A., 2014. "Klasifikasi Metode Naive Bayes Dalam Data Mining Untuk Menentukan Konsentrasi Siswa (Studi Kasus Di MAS PAB 2 Medan)". Konferensi Nasional pengembangan Teknologi Informasi dan Komunikasi (KeTIK).
- [9] Sharly & A.A, Muin. 2016. "Metode Naive Bayes Untuk Prediksi Kelulusan (Studi Kasus: Data Mahasiswa Baru Perguruan Tinggi)". Jurnal Ilmiah Ilmu Komputer, Vol. 2, No. 1

- [10] Widaningsih, Sri. 2019.” Perbandingan Metode Data Mining Untuk Prediksi Nilai Dan Waktu Kelulusan Mahasiswa Prodi Teknik Informatika Dengan Algoritma C4.5, Naïve Bayes, Knn Dan Svm”. Jurnal Tekno Insentif, Vol 13 No 1
- [11] Daniela XHEMALI, Christopher J. HINDE and Roger G. STONE3. 2009. “*Naïve Bayes vs. Decision Trees vs. Neural Networks in the Classification of Training Web Pages*”. IJCSI International Journal of Computer Science Issues, Vol. 4, No. 1.
- [12] Fauzi, Y., dkk. (2012). “*Kelapa Sawit. Jakarta: Penebar Swadaya*”. Halaman 3, 38 – 39, 178
- [13] Shwartz, Shalev and David, Shai Ben. 2014 “*Understanding Machine Learning: From Theory to Algorithms*”. Cambridge University Press.
- [14] Nugroho, A. dan Subanar. 2013. “Klasifikasi Naïve Bayes untuk Prediksi Kelahiran pada Data Ibu Hamil”. Berkala MIPA 23. 3, 297-308.
- [15] R. Siringoringo, 2018. “Klasifikasi Data Tidak Seimbang Menggunakan Algoritma SMOTE dan K-Nearest Neighbor,” J. ISD, vol. 3, no. 1, pp. 44–49
- [16] Hairani, N. A. Setiawan, & T. B. Adji. 2013. Metode Klasifikasi Data Mining dan Teknik Sampling Smote. Seminar Nasional Sains dan Teknologi, 168–172.