

**KLASIFIKASI NANAS DENGAN MENGGUNAKAN ALGORITMA
NAÏVE BAYES CLASSIFICATION PADA BALAI PENYULUHAN
PERTANIAN (BPP) PANAI TENGAH**

SKRIPSI



RAMADIANI BR RAMBE

NIM. 0703172047

PROGRAM STUDI MATEMATIKA

FAKULTAS SAINS DAN TEKNOLOGI

UNIVERSITAS ISLAM NEGERI SUMATERA UTARA MEDAN

MEDAN

2022

**KLASIFIKASI NANAS DENGAN MENGGUNAKAN ALGORITMA
NAÏVE BAYES CLASSIFICATION PADA BALAI PENYULUHAN
PERTANIAN (BPP) PANAI TENGAH**

SKRIPSI

(Diajukan Untuk Melengkapi Tugas-tugas dan Memenuhi Syarat-syarat Guna
Mendapatkan Gelar Sarjana S1 dalam Sains dan Teknologi)



RAMADIANI BR RAMBE

NIM. 0703172047

PROGRAM STUDI MATEMATIKA

FAKULTAS SAINS DAN TEKNOLOGI

UNIVERSITAS ISLAM NEGERI SUMATERA UTARA MEDAN

MEDAN

2022



KEMENTERIAN AGAMA REPUBLIK INDONESIA
UNIVERSITAS ISLAM NEGERI SUMATERA UTARA MEDAN
FAKULTAS SAINS DAN TEKNOLOGI
Jl. Willièm Iskandar Pasar V Medan Estate 20371
Telp. (061) 6615683-6622925 Fax. 6615683

PENGESAHAN SKRIPSI

Nomor: B.060/ST/ST.V.2/PP.01.1/03/2022

Judul Skripsi : Klasifikasi Nanas Dengan Menggunakan Algoritma *Naïve Bayes Classification* Pada Balai Penyuluhan Pertanian (BPP) Panai Tengah.

Nama : Ramadiani Br Rambe

NIM : 0703172047


Program Studi : Matematika

Telah dipertahankan di hadapan Dewan Penguji Skripsi Program Studi Matematika Fakultas Sains Dan Teknologi UIN Sumatera Utara Medan dan dinyatakan **LULUS**.

Pada hari/tanggal : Rabu, 23 Februari 2022

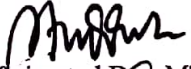
Tempat : Gedung FST UINSU

Tim Ujian Munaqasyah,
Ketua,

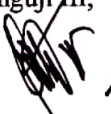

Dr. Riri Syafril Lubis, M.Si
NIDN. 2013078401

Dewan Penguji,


Penguji I,


Dr. Sajatud Dur, MT
NIDN. 2013107302

Penguji III,


Dr. Sutarman, M.Sc
NIDN. 0026106305

Penguji II,


Dr. Ismail Husein, M.Si
NIDN. 2022049101

Penguji IV


Hendra Cipta, M.Si
NIDN. 2002078902

Mengesahkan,
Dekan Fakultas Sains dan Teknologi
UIN Sumatera Utara Medan,

Dr. Mhd. Syahnan, MA
NIP. 196609051991031002

PERSETUJUAN SKRIPSI

Hal : Surat Persetujuan Skripsi
Lamp : -

Kepada Yth.,
Dekan Fakultas Sains dan Teknologi
Universitas Islam Negeri Sumatera Utara Medan

Assalamu'alaikum Wr. Wb.

Setelah membaca, meneliti, memberikan petunjuk, dan mengoreksi serta mengadakan perbaikan, maka kami selaku pembimbing berpendapat bahwa skripsi saudara:

Nama : Ramadiani Br Rambe
NIM : 0703172047
Program Studi : Matematika
Judul : Klasifikasi Nanas Dengan Menggunakan
Algoritma *Naïve Bayes Classification* Pada
Balai Penyuluhan Pertanian (BPP) Panai Tengah

Dapat disetujui untuk segera *dimunaqasyahkan*. atas perhatiannya kami ucapkan terimakasih.

Medan, Februari 2022
Komisi Pembimbing,

Pembimbing I



Dr. Sajaratud Dur, M.T.
NIDN. 2013107302

Pembimbing II



Dr. Ismail Husein, M.Si
NIDN. 2022049101

Mengetahui,
Ketua Program Studi Matematika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Sumatera Utara Medan



Dr. Riri Syafitri Lubis, S.Pd, M.Si
NIDN. 2013078401

SURAT PERNYATAAN KEASLIAN SKRIPSI

Saya yang bertanda tangan di bawah ini,

Nama : Ramadiani Br Rambe
NIM : 0703172047
Program Studi : Matematika
Judul : Klasifikasi Nanas Dengan Menggunakan Algoritma *Naïve Bayes Classification* Pada Balai Penyuluhan Pertanian (BPP) Panai Tengah

Menyatakan bahwa skripsi ini adalah hasil karya saya sendiri, kecuali beberapa kutipan dan ringkasan yang masing-masing disebutkan sumbernya. Apabila di kemudian hari ditemukan plagiat dalam skripsi ini maka saya bersedia menerima sanksi pencabutan gelar akademik yang saya peroleh dan sanksi lainnya sesuai dengan peraturan yang berlaku.

Medan, Februari 2022



Ramadiani Br Rambe
NIM. 0703172047

ABSTRAK

Tujuan dari penelitian ini adalah untuk mendapatkan hasil dari klasifikasi nanas dengan menggunakan algoritma *naïve bayes classification* pada Balai Penyuluhan Pertanian (BPP) Panai Tengah. Penelitian ini merupakan penelitian deskriptif, metode kuantitatif, dan menggunakan data sekunder yaitu data yang diperoleh melalui Balai Penyuluhan Pertanian (BPP) Panai Tengah. Adapun jenis data yang dibutuhkan meliputi karakter agronomi nanas, yaitu: umur panen (hari), daerah tanam (ha), diameter buah (mm), bobot buah (kg), panjang buah (cm), tinggi tanaman (cm), dan produksi nanas dalam satu periode tanam (kg).

Penelitian ini menggunakan metode algoritma *naïve bayes classification* karena merupakan metode yang baik di dalam mesin pembelajaran berdasarkan *data training* dengan menggunakan probabilitas bersyarat sebagai dasarnya. Selain itu *naïve bayes classification* juga memiliki kecepatan yang sangat cepat, ketepatan yang baik untuk semua data dan transparansi yang tanpa aturan. Sehingga dapat memudahkan peneliti menghitung dan menentukan kemungkinan-kemungkinan jenis nanas yang produktif dan tidak produktif dengan menggunakan semua atribut yang ada.

Hasil dari penelitian ini memperoleh nilai *accuracy* sebesar 85,29%. Untuk kelas “Ya” pada *prediction* sebesar 85,71% dan untuk kelas “Tidak” pada *prediction* sebesar 84,62%. Karena memiliki nilai lebih dari 50% ini berarti kinerja *naïve bayes* yang diterapkan dalam permasalahan penelitian ini sudah cukup baik.

Kata Kunci: Nanas Panai Tengah, Algoritma *Naïve Bayes Classification*, *Accuracy*

ABSTRACT

The purpose of this study was to obtain results from pineapple classification using the naive Bayes classification algorithm at the Central Panai Agricultural Extension Center (BPP). This research is a descriptive research, quantitative method, and uses secondary data, namely data obtained through the Center for Agricultural Extension (BPP) Panai Tengah. The types of data needed include pineapple agronomic characters, namely: harvest age (days), planting area (ha), fruit diameter (mm), fruit weight (kg), fruit length (cm), plant height (cm), and production. pineapple in one planting period (kg).

This study uses the naive Bayes classification algorithm because it is a good method in machine learning based on training data using conditional probability as the basis. In addition, naive bayes classification also has a very fast speed, good accuracy for all data and transparency without rules. So that it can make it easier for researchers to calculate and determine the possibilities of productive and unproductive pineapple types by using all existing attributes.

The results of this study obtained an accuracy value of 85.29%. For class "Yes" at prediction is 85.71% and for class "No" at prediction is 84.62%. Because it has a value of more than 50%, it means that the naive Bayes performance applied in this research problem is quite good.

Keywords: *Pineapple Panai Tengah, Naïve Bayes Classification Algorithm, Accuracy*

KATA PENGANTAR

Assalamu'alaikum Wr. Wb

Alhamdulillah, puji dan syukur kepada Allah SWT atas segala rahmat dan nikmat-Nya sehingga peneliti dapat menyelesaikan skripsi ini dengan baik. Shalawat dan salam senantiasa tetap terlimpahkan kepada Nabi kita yaitu Nabi Muhammad Saw. Semoga dengan memperbanyak mengucapkan shalawat dan salam, kita menjadi umatnya yang akan mendapatkan syafaat-Nya.

Penelitian skripsi ini ditujukan untuk memenuhi syarat memperoleh gelar S 1 Program Studi Matematika Fakultas Sains dan Teknologi Universitas Islam Negeri Sumatera Utara dengan judul “Klasifikasi Nanas Dengan Menggunakan Algoritma *Naïve Bayes Classification* Pada Balai Penyuluhan Pertanian (BPP) Panai Tengah”.

Skripsi ini dapat selesai dengan baik berkat bimbingan dan dukungan dari berbagai pihak baik materi, moral, maupun do'a, maka diucapkan terima kasih dan pemberian penghargaan yang setinggi-tingginya, kepada:

1. Ayahanda Khairul Ahmad Rambe dan Ibunda Mariani Br Ginting. Terimakasih untuk semua kasih sayang, cinta, pengorbanan, kepercayaan, dukungan dan do'a yang tidak pernah putus untuk anaknya, serta telah memberikan segala bantuan dalam bentuk materil dan moril sehingga peneliti bisa menyelesaikan perkuliahan dan sampai selesainya penelitian skripsi ini.
2. Bapak Prof. Dr. Syahrin Harahap, M.A. selaku Rektor Universitas Islam Negeri Sumatera Utara.
3. Bapak Dr. Mhd. Syahnan, M.A. selaku Dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Sumatera Utara.
4. Ibu Dr. Riri Syafitri Lubis, S.Pd, M.Si. selaku Ketua Prodi Matematika Fakultas Sains dan Teknologi Universitas Islam Negeri Sumatera Utara Medan sekaligus Dosen Pembimbing Akademik.

5. Ibu Rima Aprilia, M.Si. selaku Sekretaris Prodi Matematika Fakultas Sains dan Teknologi Universitas Islam Negeri Sumatera Utara Medan yang telah banyak memberikan bimbingan dan arahan kepada ananda, serta kepada seluruh staff Prodi Matematika yang menjadi jembatan ananda sampai pada tahap ini.
6. Dr. Sajaratud Dur, M.T. dan Dr. Ismail Husein, M.Si. selaku pembimbing skripsi yang telah berjasa dan sangat sabar dalam membimbing peneliti dalam menyelesaikan tugas akhir ini.
7. Kepada seluruh teman-teman jurusan Matematika stambuk 2017 yang telah mendukung dan memberikan semangat kepada peneliti dalam menyelesaikan skripsi ini.

Banyak kekurangan dan kelemahan dalam penelitian ini. Oleh karena itu peneliti mengharapkan kritik dan saran yang sifatnya membangun. Semoga skripsi ini bermanfaat dalam menambah wawasan keilmuan di Fakultas Sains dan Teknologi UIN Sumatera Utara Medan. Aamiin.

Medan, Februari 2022

Peneliti

Ramadiani Br Rambe

NIM. 0703172047

DAFTAR ISI

COVER SKRIPSI

PERSETUJUAN SKRIPSI

PENGESAHAN SKRIPSI

SURAT PERNYATAAN KEASLIAN SKRIPSI

ABSTRAK i

ABSTRACT ii

KATA PENGANTAR..... iii

DAFTAR ISI.....v

DAFTAR GAMBAR..... viii

DAFTAR TABEL ix

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah 1

1.2 Rumusan Masalah.....4

1.3 Batasan Masalah4

1.4 Tujuan Penelitian5

1.5 Manfaat Penelitian5

BAB II

TINJAUAN PUSTAKA

2.1 *Data Mining*6

2.1.1 Model Data Mining6

2.1.2 Klasifikasi.....6

2.2 *Naïve Bayes Classification*7

2.2.1 Probabilitas *Prior*7

2.2.2	Peluang Bersyarat (<i>Likelihood</i>)	8
2.2.3	<i>Continuous</i> dan Eksposisi Berkelanjutan	9
2.2.4	Probabilitas <i>Posterior</i>	10
2.2.5	<i>Evidence</i>	11
2.2.6	Estimasi Maksimum <i>Likelihood</i>	12
2.3	Matriks	12
2.3.1	<i>Confusion Matrixt</i>	12
2.3.2	<i>Accuracy</i>	13
2.3.3	<i>Precision</i> dan <i>Recall</i>	13
2.3.4	<i>False Positives</i>	14
2.3.5	<i>Sentiment Error</i>	14
2.4	Nanas	15
2.4.1	Manfaat Nanas	16
2.5	Wahdatul Ulum	16
2.6	Penelitian Terdahulu	18

BAB III

METODE PENELITIAN

3.1	Tempat dan Waktu Penelitian	21
3.2	Jenis Penelitian	21
3.3	Variabel Penelitian	21
3.4	Prosedur Penelitian	22
3.5	Diagram Alir	23

BAB IV

HASIL DAN PEMBAHASAN

4.1	Pengumpulan Data	24
-----	------------------------	----

4.2	Baca <i>Data Training</i> Tabel 4.1	26
4.3	Menghitung Probabilitas <i>Prior</i>	26
4.4	Menghitung Nilai <i>Likelihood</i>	26
4.4.1	Menghitung Nilai <i>Mean</i> dan Standar Deviasi	27
4.4.2	Perhitungan prediksi dengan <i>Naïve Bayes</i> dengan Fungsi Dentitas Gaus	38
4.4.3	Menghitung Probabilitas <i>Posterior</i>	43
4.4.4	Mencari Nilai <i>Likelihood</i> Maksimum	43
4.5	Analisa Algoritma <i>Naïve Bayes</i> dengan <i>RapidMiner 7.1</i>	44
4.5.1	Proses Split Data	44
4.5.2	Akurasi Prediksi	46
4.5.3	Kurva ROC/AUC (<i>Area Under Curve</i>)	48
4.6	Pembahasan.....	49
 BAB V		
KESIMPULAN DAN SARAN		
5.1	Kesimpulan	51
5.2	Saran	51
DAFTAR PUSTAKA		52
LAMPIRAN.....		54

DAFTAR GAMBAR

Gambar 3.2 Diagram Alir	31
Gambar 4.1 Operator Read Excel dan Split Data	44
Gambar 4.2 Alur Split Data	45
Gambar 4.3 ExampleSet Split Data	45
Gambar 4.4 Susunan Operator Read Excel, Split Data, Naïve Bayes, Apply Model, dan Performance	46
Gambar 4.5 Icon Run	47
Gambar 4.6 Performance Accuracy	47
Gambar 4.7 Performance Precision.....	48
Gambar 4.8 Performance Recall	48
Gambar 4.9 Curva ROC.....	49

DAFTAR TABEL

Tabel 4.1 <i>Data Training</i>	25
Tabel 4.2 Probabilitas <i>Prior</i>	26

BAB 1

PENDAHULUAN

1.1 Latar Belakang Masalah

Data Mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai basis data besar (Saputro & Sari, 2019: 2). Menurut (Budiyantara et al., 2022: 3) Ada beberapa macam teknik data mining, namun ada 3 yang paling populer antara lain: *association rules*, *clustering*, *classification*.

Association rules adalah metode yang digunakan untuk mengidentifikasi beberapa hubungan yang menarik antara variabel yang berbeda di dalam *database* yang besar, *classification* merupakan teknik yang digunakan untuk mengambil informasi penting dan relevan tentang data dan juga metadata dan dalam klasifikasi terdapat target variabel kategori. Sebagai contoh, penggolongan pendapatan dapat dipisahkan dalam 3 kategori, yaitu pendapatan tinggi, pendapatan sedang dan pendapatan rendah. Sedangkan *clustering* merupakan pengelompokan *record*, pengamatan atau memperhatikan dan membentuk kelas objek-objek yang memiliki kemiripan. Kluster adalah kumpulan *record* yang memiliki kemiripan satu dengan yang lainnya dan memiliki ketidakmiripan dengan *record-record* dalam kluster lain.

Pengklasteran berbeda dengan klasifikasi yaitu tidak adanya variabel target dalam pengklasteran. Pengklasteran tidak mencoba untuk melakukan klasifikasi, mengestimasi, memprediksi nilai dari variabel target. Akan tetapi, algoritma pengklasteran mencoba untuk melakukan pembagian terhadap keseluruhan data menjadi kelompok-kelompok yang memiliki kemiripan (homogen), yang mana kemiripan *record* dalam satu kelompok akan bernilai maksimal, sedangkan kemiripan dengan *record* dalam kelompok lain akan bernilai minimal (Defiyanti, 2017: 257). Menurut (Luvia et al., 2017: 76) Perbedaan *classification* dengan metode *clustering* terletak pada data, di mana pada *clustering* variabel dependen tidak ada, sedangkan pada *classification* diharuskan ada variabel dependen. Sehingga dalam penelitian ini peneliti menggunakan model data *mining classification*.

Pada data mining terdiri dari banyak algoritma yang dapat digunakan untuk proses klasifikasi yaitu *Naive Bayes*, *Decision Tree*, *K-Nearest Neighbor* (KNN), *Neural Network* (Ayudhitama & Pujianto, 2020: 2). *Naive bayes classifier* atau disebut juga dengan bayesian *classification* merupakan metode pengklasifikasian statistik yang didasarkan pada teorema *bayes* yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu kelas. *Bayesian classification* terbukti memiliki akurasi dan kecepatan yang tinggi saat diaplikasikan ke dalam database yang besar (Muhamad et al., 2017: 181).

Decision tree merupakan salah satu metode klasifikasi yang menggunakan representasi struktur pohon (*tree*) (Partogi, 2021: 21). *Decision tree* ini data yang berupa fakta dirubah menjadi sebuah pohon keputusan yang berisi aturan dan tentunya dapat lebih mudah dipahami dengan bahasa alami (Verawati & Hasibuan, 2022: 159). Sedangkan *k-nearest neighbor* (*K-NN*) adalah algoritma pengklasifikasian yang didasarkan pada analogi, yaitu membandingkan data uji dengan data pelatihan yang berada dekat dan memiliki kemiripan dengan data uji tersebut (Sartika & Indra, 2017: 153), dan *neural network* merupakan satu set unit input ataupun output yang terhubung yang mana setiap relasinya memiliki bobot (Yudhistira et al., 2022: 33).

Diantara keempat algoritma pengklasifikasian tersebut yang lebih fleksibel yaitu *naive bayes* karena memiliki akurasi dan nilai AUC yang paling tinggi (Kadafi, 2018). *Naive bayes* hanya memerlukan sejumlah kecil data pelatihan untuk mengestimasi parameter yang dibutuhkan untuk klasifikasi. Selain itu, *naive bayes* juga mampu menangani nilai yang hilang dengan mengabaikan instansi selama perhitungan estimasi peluang (Mayadewi & Rosely, 2015: 330).

Metode *naive bayes classification* dipilih karena merupakan metode yang baik di dalam mesin pembelajaran berdasarkan data *training* dengan menggunakan probabilitas bersyarat sebagai dasarnya. Selain itu *naive bayes classification* juga memiliki kecepatan yang sangat cepat, ketepatan yang baik untuk semua data dan transparansi yang tanpa aturan. *Naive bayes* adalah metode *bayesian learning* yang paling cepat dan sederhana. Hal ini berasal dari teorema bayes dan hipotesis kebebasan, menghasilkan klasifikasi statistik berdasarkan peluang (Suryono et al., 2017: 40).

Berdasarkan data BPS Labuhanbatu tahun 2020 produksi nanas mengalami penurunan. Produksi nanas pada tahun 2020 sebanyak 109 ton, sementara di tahun 2019 produksi nanas sebanyak 325 ton. Terlihat di sini bahwa pengembangan nanas kurang mendapat perhatian serius dari pemerintah Provinsi Sumatera Utara, mengingat belum berkembangnya penggunaan varietas unggul dan belum optimalnya teknik pembudidayaan tanaman nanas.

Menurut data BPS Labuhanbatu bahwa terdapat 2 Kecamatan di Kabupaten Labuhanbatu yang melakukan budidaya nanas, yaitu: Kecamatan Bilah Hilir dan Kecamatan Panai Tengah. Dari kedua Kecamatan tersebut Kecamatan Panai Tengah sangat berpotensi. Luas lahan nanas di Panai Tengah awalnya mencapai 300 ha, saat ini hanya berkisar 100 ha dan terdapat 86 Petani. Kecamatan Panai Tengah petani nanas memanfaatkan lahan gambut untuk budidaya nanas, lahan gambut di Panai Tengah merupakan gambut garaman di mana Panai Tengah berada di wilayah pesisir pantai. Sehingga para petani membudidayakan nanas dengan potensi produksi tinggi dibandingkan dengan lahan lainnya.

Berkurangnya petani nanas di Panai Tengah akibat banyak petani beralih cocok tanam ke tanaman perkebunan, karena petani mengalami kesulitan dalam perawatan buah nanas. Pertumbuhan produksi nanas yang lebih rendah dari pada luas panennya menyebabkan produktivitas nanas mengalami penurunan. Maka, untuk mendukung program swasembada nanas di Panai Tengah, tahap awal yang dilakukan adalah meningkatkan produksi dan produktivitas nanas.

Klasifikasi pernah dilakukan oleh beberapa peneliti diantaranya Hadit-sah Annur (2018) dengan menggunakan algoritma *naive bayes* untuk meneliti masyarakat miskin terkait dengan adanya pertumbuhan ekonomi yang tidak tersebar secara merata di Kecamatan Tibawa Kabupaten Gorontalo, Fakhriani Ekawati (2018) meneliti penentuan jurusan pada siswa madrasah Aliyah menggunakan algoritma *naive bayes*, dan Tiara Imandasari, Dkk (2019) meneliti klasifikasi lokasi pembangunan sumber air menggunakan algoritma *naive bayes*.

Berdasarkan latar belakang di atas maka dilakukan penelitian dengan judul klasifikasi nanas menggunakan algoritma *naive bayes classification* pada Balai Penyuluhan Pertanian (BPP) Panai Tengah. Dalam penelitian ini algoritma *naive bayes classification* dipilih karena algoritma *naive bayes classification* dapat memudahkan peneliti menghitung dan menentukan kemungkinan-kemungkinan jenis nanas yang produktif dan tidak produktif dengan menggunakan semua atribut yang ada.

1.2 Rumusan Masalah

K-Nearest Neighbor adalah sebuah metode *supervised* yang berarti membutuhkan data *training* untuk mengklasifikasikan objek yang jaraknya paling dekat. Prinsip kerja *k-nearest neighbor* adalah mencari jarak terdekat antara data yang akan di evaluasi dengan k tetangga (*neighbor*) dalam data pelatihan. Sedangkan *naive bayes* merupakan metode *supervised document classification* yang berarti membutuhkan data *training* sebelum melakukan proses klasifikasi. Algoritma *naive bayes classification* sangat cocok diterapkan pada pengklasifikasian yang memiliki banyak atribut serta dapat menangani nilai yang hilang. Demikian akan dilakukan pengklasifikasian nanas dengan menggunakan algoritma *naive bayes classification* pada Balai Penyuluhan Pertanian (BPP) Panai Tengah.

1.3 Batasan Masalah

Dalam penelitian ini adanya batasan-batasan masalah agar tujuan penelitian dapat tercapai, yaitu:

- a. Penelitian ini hanya mengklasifikasi data yang diperoleh dari instansi Balai Penyuluhan Pertanian (BPP) Kecamatan Panai Tengah, Kabupaten Labuhanbatu, Provinsi Sumatera Utara.
- b. Metode yang dipakai dalam penelitian ini adalah algoritma *naive bayes classification*.
- c. Atribut yang digunakan dalam perhitungan *naive bayes* adalah umur panen, daerah tanam, diameter buah, bobot buah, panjang buah, tinggi tanaman, jumlah produksi nanas dalam satu periode tanam.

1.4 Tujuan Penelitian

Berdasarkan dari rumusan masalah yang ada maka tujuan dari penelitian ini adalah untuk mendapatkan hasil dari klasifikasi nanas dengan menggunakan algoritma *naive bayes classification* pada Balai Penyuluhan Pertanian (BPP) Panai Tengah.

1.5 Manfaat Penelitian

Adapun manfaat yang dapat diambil dari penelitian yang dilakukan adalah:

- a. Dapat menghasilkan informasi dalam mengetahui hasil kinerja *naive bayes classification* yang diterapkan dalam klasifikasi nanas di Balai Penyuluhan Pertanian (BPP) Panai Tengah.
- b. Dapat mengembangkan wawasan keilmuan yang luas, menjadi sarana pengembangan berfikir ilmiah dan rasional, serta diharapkan dapat diimplementasikan.
- c. Sebagai bahan acuan peneliti selanjutnya dalam mengembangkan penelitian yang lebih lanjut.

BAB 2

TINJAUAN PUSTAKA

2.1 *Data Mining*

Data mining merupakan teknik yang dapat membantu memprediksi fenomena suatu objek yang diteliti dengan terlebih dahulu mengetahui atau tidak mengetahui suatu kelas suatu obyek tertentu melalui masukan dari data besar (Natasuwarna, 2019: 24).

Data mining adalah proses maupun tahapan dalam menemukan sebuah struktur data. Struktur data tersebut dapat mengambil banyak bentuk, termasuk aturan, grafik atau jaringan, pohon (*tree*) maupun persamaan, serta beberapa yang lain. Dengan menggunakan *data mining*, maka sebuah kasus dapat dilihat *trend*, struktur maupun prediksinya di masa mendatang. *Data mining* sendiri memiliki banyak tahapan dan teknik yang dapat diimplementasikan dalam kehidupan nyata (Kurniawan, 2018: 456).

2.1.1 Model *Data Mining*

Ada berbagai model dalam *data mining* atau sering disebut teknik *data mining*, secara umum model *data mining* dibagi dalam tiga kelompok berdasarkan pada tugas atau fungsi yang terdiri dari *classification*, *clustering*, *association*. Dalam penelitian ini peneliti menggunakan model *data mining* klasifikasi (Nofi, 2018: 14).

2.1.2 Klasifikasi

Klasifikasi adalah teknik pengolahan data yang membagi objek menjadi beberapa kelas sesuai dengan jumlah kelas yang diinginkan (Arifin & Fitri-anah, 2018). Cara kerja dari metode *classification* adalah sebuah proses 2 langkah. Langkah pertama dari *classification* juga disebut sebagai *learning of mapping* atau *function*, suatu fungsi pemetaan yang bisa memprediksi *class label Y* pada suatu *tuple X*. Pemetaan ini direpresentasikan dalam bentuk *classification rules*, *decision tree* atau formula matematika. Dari *rules* atau *tree* tersebut dapat digunakan untuk mengklasifikasi *tuple* baru. Langkah kedua adalah *classifier* yang sudah dibangun akan digunakan untuk mengklasifikasi data. Pertama, akurasi dari prediksi *classifier* tersebut diperkirakan. Jika

menggunakan *training set* untuk mengukur akurasi dari *classifier*, maka estimasi akan optimis karena data yang digunakan untuk membentuk *classifier* adalah *training set* juga. Oleh karena itu, digunakan *test set*, yaitu sekumpulan *tuple* beserta *class label*-nya yang dipilih secara acak dari dataset. *Test set* bersifat independen dari *training set* dikarenakan *test set* tidak digunakan untuk membangun *classifier* (Atika & Priatna, 2020: 89).

2.2 Naive Bayes Classification

Naive bayes classification adalah pengklasifikasi linier yang dikenal sederhana namun sangat efisien. Model probabilistik pengklasifikasi *naive bayes* didasarkan pada teorema *bayes*, dan kata sifat *naive* berasal dari asumsi bahwa fitur dalam dataset saling independen. Dalam praktiknya, asumsi independensi sering dilanggar, tetapi pengklasifikasi *naive bayes* masih cenderung berkinerja sangat baik di bawah asumsi yang tidak realistis ini. Khusus untuk ukuran sampel kecil, pengklasifikasi *naive bayes* dapat mengungguli alternatif yang lebih kuat.

Menjadi relatif kuat, mudah diimplementasikan, cepat, dan akurat, pengklasifikasi *naive bayes* digunakan di berbagai bidang. Beberapa contoh termasuk diagnosis penyakit dan membuat keputusan tentang proses pengobatan, klasifikasi urutan RNA dalam studi taksonomi, dan penyaringan spam di *klien email*. Namun, pelanggaran kuat dari asumsi independensi dan masalah klasifikasi *non-linier* dapat menyebabkan kinerja pengklasifikasi *naive bayes* yang sangat buruk. Kita harus ingat bahwa tipe data dan masalah tipe yang akan dipecahkan menentukan model klasifikasi mana yang ingin kita pilih. Dalam praktiknya, selalu disarankan untuk membandingkan model klasifikasi yang berbeda pada kumpulan data tertentu dan mempertimbangkan kinerja prediksi serta efisiensi komputasi.

2.2.1 Probabilitas *Prior*

Berbeda dengan pendekatan *frequentist*, tambahan probabilitas *prior* (atau hanya *prior*) diperkenalkan dapat diartikan sebagai kepercayaan sebelumnya atau sebuah prioritas pengetahuan.

Dalam konteks klasifikasi pola, probabilitas *prior* juga disebut *class prior*, yang menggambarkan “probabilitas umum untuk bertemu dengan kelas tertentu”. Jika *prior* mengikuti distribusi seragam, probabilitas *posterior* akan sepenuhnya ditentukan oleh probabilitas bersyarat dan istilah *evidence*, dan karena istilah *evidence* adalah konstan, aturan keputusan akan sepenuhnya bergantung pada probabilitas bersyarat (mirip dengan pendekatan frequentist dan estimasi maksimum *likelihood*).

Selanjutnya, sebuah prioritas pengetahuan dapat diperoleh, misalnya dengan berkonsultasi dengan pakar domain atau dengan estimasi dari *data training* (dengan asumsi bahwa *data training* adalah data *i.i.d.* Singkatan dari “*independent and identically distributed*” dan sampel yang mewakili seluruh populasi. pendekatan estimasi maksimum *likelihood* dapat dirumuskan sebagai:

$$\hat{P}(w_j) = \frac{N_{w_j}}{N_c} \quad (2.1)$$

N_{w_j} = Jumlah sampel dari kelas w_j

N_c = Hitungan semua sampel

2.2.2 Peluang Bersyarat (*Likelihood*)

Satu asumsi yang dibuat oleh *bayes classifier* adalah bahwa sampelnya yaitu *i.i.d.* Menggambarkan variabel acak yang independen satu sama lain dan diambil dari distribusi probabilitas yang serupa. Independensi berarti bahwa probabilitas satu pengamatan tidak mempengaruhi probabilitas pengamatan lain (misalnya, deret waktu dan grafik jaringan tidak independen). Salah satu contoh populer dari *i.i.d.* variabel adalah lemparan koin klasik: Lemparan koin pertama tidak mempengaruhi hasil lemparan koin kedua dan seterusnya. Diberikan koin yang adil, kemungkinan koin mendarat di “gambar” selalu 0,5 tidak peduli seberapa sering koin itu jika dibalik.

Asumsi tambahan dari *naive bayes classifier* adalah fitur bebas bersyarat. Asumsi *naive*, itu probabilitas kelas bersyarat atau (*likelihood*) dari sampel dapat diperkirakan langsung dari *data training* sebagai gantinya mengevaluasi semua kemungkinan x . Jadi, diberikan D-dimensi fitur vektor x , probabilitas kelas bersyarat dapat dihitung sebagai berikut:

$$P(x|w_j) = P(x_1|w_j) \cdot P(x_2|w_j) \dots P(x_d|w_j) = \prod_{k=1}^d P(x_k|w_j) \quad (2.2)$$

Di Sini, $P(x|w_j)$ secara sederhana berarti: “Seberapa besar kemungkinan untuk mengamati pola x mengingat itu milik kelas w_j ?” Kemungkinan “individual” *likelihood* untuk setiap fitur dalam vektor fitur dapat diperkirakan melalui estimasi kemungkinan maksimum, yang hanya merupakan frekuensi dalam kasus data kategorikal:

$$\hat{P}(x_i|w_j) = \frac{N_{x_i, w_j}}{N_{w_j}} \quad (i = 1, \dots, d) \quad (2.3)$$

N_{x_i, w_j} = Berapa kali fitur x_i muncul dalam sampel dari kelas w_j

N_{w_j} = Jumlah total semua fitur di kelas w_j

2.2.3 *Continuous* dan Eksposisi Berkelanjutan

Dalam pendekatan *Bayesian*, istilah “*prior*”, “*likelihood*”, dan “*posterior*” biasanya digunakan dan kami mengeksplorasi terminologi ini di sini. Kami biasa tertarik pada parameter θ , rata-rata distribusi beberapa data x (saya menggunakan notasi standar di sini). Tetapi dalam pengaturan bayesian kita tidak hanya menginginkan nilai θ , tetapi kita menginginkan distribusi nilai θ yang dimulai dari beberapa asumsi sebelumnya tentang distribusi ini. Jadi kita mulai dengan $p(\theta)$, yang kita sebut distribusi *prior*. Selanjutnya kami mengamati data x , dan menggabungkan data dengan sebelumnya untuk mendapatkan distribusi *posterior* $p(\theta|x)$. Untuk melakukan ini, kita perlu menghitung probabilitas data x dengan $p(\theta)$ sebelumnya dan probabilitas ini diberikan oleh fungsi *likelihood* $L(x|\theta)$. Asumsikan bahwa varians dari data x diketahui, yaitu σ^2 .

Menerapkan teorema bayes yang kita miliki

$$p(\theta|x) = \frac{L(x|\theta) p(\theta)}{\int L(x|\theta) p(\theta) d\theta} \propto L(x|\theta) p(\theta) \quad (2.4)$$

Jika kita mengasumsikan distribusi *prior* untuk mean data adalah normal, yaitu $p(\theta) \sim N[\mu_0, \sigma_0^2]$ dan *likelihood* juga normal, yaitu $L(x|\theta) \sim N[\theta, \sigma^2]$ maka kita memiliki

$$\begin{aligned}
p(\theta) &= \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left[-\frac{1}{2}\frac{(\theta-\mu_0)^2}{\sigma_0^2}\right] \sim N[\theta|\mu_0, \sigma_0^2] \propto \exp\left[-\frac{1}{2}\frac{(\theta-\mu_0)^2}{\sigma_0^2}\right] \\
L(x|\theta) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\frac{(x-\theta)^2}{\sigma^2}\right] \sim N[x|\theta, \sigma^2] \propto \exp\left[-\frac{1}{2}\frac{(x-\theta)^2}{\sigma^2}\right]
\end{aligned} \tag{2.5}$$

Kemudian *posterior* mengikuti, sebagai berikut:

$$p(\theta|x) \propto L(x|\theta) p(\theta) \propto \exp\left[-\frac{1}{2}\frac{(x-\theta)^2}{\sigma^2} - \frac{1}{2}\frac{(\theta-\mu_0)^2}{\sigma_0^2}\right] \tag{2.6}$$

Tentukan nilai presisi menjadi $\tau_0 = 1/2\sigma_0^2$, dan $\tau = 1/2\sigma^2$. Kemudian dapat ditunjukkan bahwa ketika Anda mengamati nilai baru dari data x , distribusi *posterior* ditulis dalam bentuk tertutup sebagai:

$$p(\theta|x) \sim N\left[\frac{\tau_0}{\tau_0 + \tau}\mu_0 + \frac{\tau}{\tau_0 + \tau}x, \frac{1}{\tau_0 + \tau}\right] \tag{2.7}$$

Ketika distribusi *posterior* dan distribusi *prior* memiliki bentuk yang sama, mereka dikatakan “konjugasi” sehubungan dengan fungsi *likelihood* yang spesifik. Sebagai contoh, misalkan *prior* kita untuk mean ekuitas $p(\theta) \sim N[0.005, 0.001^2]$, maka standar deviasi dari premi ekuitas adalah 0,04

Oleh karena itu, kami melihat bahwa setelah memperbarui rata-rata telah sedikit meningkat karena data yang masuk lebih tinggi dari yang diharapkan. Jika kita mengamati n nilai baru dari X , maka *posterior* yang baru adalah:

$$p(\theta|x) \sim N\left[\frac{\tau_0}{\tau_0 + n\tau}\mu_0 + \frac{\tau}{\tau_0 + n\tau}\sum_{j=1}^n x_j, \frac{1}{\tau_0 + n\tau}\right] \tag{2.8}$$

Ini mudah untuk diturunkan, karena ini hanya hasil yang anda peroleh jika anda mengambil setiap x_j dan memperbarui *posterior* satu per satu.

2.2.4 Probabilitas *Posterior*

Untuk memahami bagaimana *naive bayes classification* bekerja, kita harus secara singkat merekapitulasi konsep aturan bayes. Model probabilitas yang dirumuskan oleh Thomas Bayes (1701-1761) cukup sederhana namun kuat. Dapat dituliskan dengan kata-kata sederhana sebagai berikut:

$$posterior = \frac{likelihood \times prior}{evidence} \tag{2.9}$$

Teorema *bayes* membentuk inti dari seluruh konsep *naive bayes classification*. Probabilitas *posterior*, dalam konteks masalah klasifikasi, dapat diartikan sebagai: “Berapa peluang suatu objek tertentu masuk ke dalam kelas i untuk nilai fitur yang diamati?” Contoh yang lebih konkrit adalah: “Berapa peluang seseorang menderita diabetes?”.

Notasi umum dari probabilitas *posterior* dapat ditulis sebagai berikut:

$$P(w_j|x_i) = \frac{P(x_i|w_j) \cdot P(w_j)}{P(x_i)} \quad (2.10)$$

Fungsi tujuan dalam probabilitas *naive bayes* adalah untuk memaksimalkan probabilitas *posterior* yang diberikan *data training* untuk merumuskan aturan keputusan.

$$\text{label kelas yang diprediksi} \leftarrow \arg \max_{j=1, \dots, m} P(w_j|x_i) \quad (2.11)$$

2.2.5 Evidence

Setelah mendefinisikan probabilitas bersyarat dan probabilitas *prior*, hanya ada satu istilah yang hilang untuk dihitung probabilitas *posterior*, adalah *evidence*. *Evidence* $P(x)$ dapat dipahami sebagai kemungkinan menemukan pola tertentu x independen dari label kelas. Mengingat lebih formal definisi probabilitas *posterior*.

$$P(w_j|x_i) = \frac{P(x_i|w_j) \cdot P(w_j)}{P(x_i)} \quad (2.12)$$

Evidence dapat dihitung sebagai berikut (w_j^C singkatan dari “*complement*” dan pada dasarnya diterjemahkan menjadi “*not class w_j*”)

$$P(x_i) = P(x_i|w_j) \cdot P(w_j) + P(x_i|w_j^C) \cdot P(w_j^C) \quad (2.13)$$

Meskipun istilah *evidence* diperlukan untuk menghitung probabilitas *posterior* secara akurat, istilah ini dapat dihilangkan dari aturan keputusan “Mengklasifikasikan sampel x_i sebagai w_1 jika $P(w_1|x_i) > P(w_2|x_i)$ mengklasifikasikan sampel sebagai w_2 ”, karena ini hanyalah faktor penskalaan:

$$\frac{P(x_i|w_1) \cdot P(w_1)}{P(x_i)} > \frac{P(x_i|w_2) \cdot P(w_2)}{P(x_i)} \quad (2.14)$$

$$\alpha P(x_i|w_1) \cdot P(w_1) > P(x_i|w_2) \cdot P(w_2)$$

2.2.6 Estimasi Maksimum *Likelihood*

Aturan keputusan dapat didefinisikan sebagai:

$$\begin{aligned} & \text{Classify sample as } + \text{ if} \\ & P(w = +|x) \geq P(w = -|x), \text{ else classify sample as } - \end{aligned} \quad (2.15)$$

Dengan asumsi bahwa sampel adalah *i.i.d*, probabilitas *prior* dapat diperoleh melalui perkiraan maksimum *likelihood* yaitu, frekuensi seberapa sering setiap label kelas diwakili dalam kumpulan *data training* (Raschka, 2014: 3-10).

2.3 Matriks

Mengembangkan analitik tanpa matriks tidak cukup. Penting untuk membangun ukuran yang menguji apakah analitik menghasilkan klasifikasi yang signifikan secara statistik, berguna secara ekonomi, dan stabil. Agar analitik menjadi valid secara statistik, maka harus memenuhi beberapa kriteria yang menandakan akurasi dan kekuatan klasifikasi. Menjadi berguna secara ekonomi menetapkan standar yang berbeda, apakah itu menghasilkan uang? Dan stabilitas ada dua yaitu: satu, apakah kinerjanya baik di dalam dan di luar sampel? Dan dua, apakah perilaku algoritma stabil di seluruh korpora *training*? Di sini, kami menjelajahi beberapa matriks yang telah dikembangkan, dan mengusulkan yang lain. Tidak diragukan lagi, seiring dengan bertambahnya rentang analitik, demikian pula rentang matriks.

2.3.1 *Confusion Matrix*

Confusion matrix adalah alat klasik untuk menilai akurasi klasifikasi. Diberikan n kategori, matriksnya berdimensi $n \times n$. Baris-baris tersebut berhubungan dengan kategori yang ditetapkan oleh algoritma analitik dan kolom-kolom mengacu pada kategori yang benar di mana teks berada. Setiap *cell* (i, j) dari matriks berisi jumlah pesan teks yang bertipe j dan diklasifikasikan sebagai tipe i . *Cell-cell* pada diagonal dari *confusion matrix* menyatakan berapa kali algoritma mendapatkan klasifikasi yang benar. Semua *cell* lain adalah contoh kesalahan klasifikasi. Jika suatu algoritma tidak memiliki kemampuan klasifikasi, maka baris dan kolom matriks akan saling bebas. Berdasarkan hipotesis nol ini, statistik yang diuji untuk penolakan adalah sebagai berikut:

$$x^2 [dof = (n - 1)^2] = \sum_{i=1}^n \sum_{j=1}^n \frac{[A(i, j) - E(i, j)]^2}{E(i, j)} \quad (2.16)$$

Di mana $A(i, j)$ adalah bilangan aktual yang diamati dalam *confusion matrix*, dan $E(i, j)$ adalah bilangan yang diharapkan, dengan asumsi tidak ada kemampuan klasifikasi di bawah nol. Jika $T(i)$ mewakili total seluruh baris i dari *confusion matrix*, dan $T(j)$ total kolom, maka:

$$E(i, j) = \frac{T(i) \times T(j)}{\sum_{i=1}^n T(i)} \equiv \frac{T(i) \times T(j)}{\sum_{j=1}^n T(j)} \quad (2.17)$$

Derajat kebebasan dari statistik x^2 adalah $(n - 1)^2$. Statistik ini sangat mudah diimplementasikan dan dapat diterapkan pada model untuk n . Statistik yang sangat signifikan adalah bukti kemampuan klasifikasi.

2.3.2 Accuracy

Akurasi algoritma pada skema klasifikasi adalah persentase teks yang diklasifikasikan dengan benar. Ini dapat dilakukan dalam sampel atau di luar sampel. Untuk menghitung ini dari *confusion matrix*, kami menghitung

$$Accuracy = \frac{\sum_{i=1}^n A(i, i)}{\sum_{j=1}^n T(j)} \quad (2.18)$$

Kita berharap bahwa ini setidaknya lebih besar dari $1/n$, yang merupakan tingkat akurasi yang dicapai rata-rata dari tebakan acak. Dalam praktiknya, saya menemukan bahwa rasio akurasi 60 - 70%, dari rasio akurasi tersebut masuk akal untuk teks yang tidak faktual dan berisi bahasa dan opini yang buruk.

2.3.3 Precision dan Recall

Penciptaan *confusion matrix* secara alami mengarah pada dua ukuran yang terkait dengannya. Presisi adalah fraksi positif yang diidentifikasi yang benar-benar positif, dan juga dikenal sebagai nilai prediksi positif. Ini adalah ukuran kegunaan prediksi. Jadi jika algoritma ditugaskan untuk memilih pemegang akun di LinkedIn yang benar-benar mencari pekerjaan, dan itu mengidentifikasi n orang yang hanya m yang benar-benar mencari pekerjaan, maka presisinya adalah m/n .

Recall adalah proporsi positif yang diidentifikasi dengan benar, dan juga dikenal sebagai sensitivitas. Ini adalah ukuran seberapa lengkap prediksi itu. Jika jumlah sebenarnya orang yang mencari pekerjaan di LinkedIn adalah M , maka penarikan akan menjadi n/M .

2.3.4 *False Positives*

Klasifikasi yang tidak tepat lebih buruk daripada kegagalan untuk mengklasifikasikan. Dalam skema 2×2 (dua kategori, $n = 2$), setiap elemen di luar diagonal dalam *confusion matrix* adalah *false positives*. Ketika $n > 2$, beberapa kesalahan klasifikasi lebih buruk daripada yang lain. Misalnya dalam skema beli, tahan, jual 3 arah, di mana kita memiliki teks saham untuk klasifikasi, mengklasifikasikan beli sebagai jual lebih buruk daripada mengklasifikasikannya sebagai penangguhan. Dalam pengertian ini, pengurutan kategori berguna agar klasifikasi yang salah ke dalam kategori dekat tidak seburuk klasifikasi yang salah ke dalam kategori jauh (berlawanan secara diametral).

False positives adalah matriks yang berguna untuk digunakan. Ini dapat dihitung sebagai penghitungan sederhana atau sebagai penghitungan tertimbang (berdasarkan kedekatan Persentase kategori yang salah) dari klasifikasi *false* dibagi dengan total klasifikasi yang dilakukan.

Dalam percobaan kami pada pesan saham di Das dan Chen (2007), kami menemukan bahwa tingkat *false positives* untuk pengklasifikasi skema pemungutan suara adalah sekitar 10%. Ini dikurangi menjadi di bawah setengah jumlah itu setelah penerapan filter ambiguitas berdasarkan Penyelidik Umum.

2.3.5 *Sentiment Error*

Ketika banyak artikel teks diklasifikasikan, ukuran keseluruhan sentimen dapat dihitung. Agregasi berguna karena memungkinkan kesalahan klasifikasi untuk dibatalkan jika pembelian disalah artikan sebagai penjualan, dan penjualan lainnya sebagai pembelian, maka indeks sentimen agregat tidak terpengaruh.

Sentiment error adalah perbedaan persentase antara sentimen agregat yang dihitung, dan nilai yang akan kita peroleh jika tidak ada kesalahan klasifikasi. Dalam percobaan kami, ini bervariasi dari 5 - 15% di seluruh kumpulan data yang kami gunakan. Leinweber dan Sisk (2010) menunjukkan bahwa agregasi sentimen memberikan hubungan yang lebih baik antara berita dan *return* saham (Das, 2016: 207-210).

2.4 Nanas

Sektor pertanian merupakan sektor yang mempunyai peranan strategis dalam struktur pembangunan perekonomian nasional, salah satunya adalah tanaman hortikultura. Menurut kementerian pertanian (2016), nanas merupakan salah satu komoditas unggul tanaman hortikultura Indonesia yang telah dikenal di seluruh dunia dan masuk dalam lima besar produsen nanas dunia. Menurut Badan Pusat Statistik (2017), produktivitas nanas di Indonesia menempati posisi ke empat dunia dengan total produksi mencapai 1.795.986 ton (Tiara et al., 2020: 81).

Nanas merupakan tanaman buah berupa semak yang memiliki nama ilmiah *Ananas comosus (L) Merr* yang memiliki bagian-bagian seperti akar, daun, buah, mahkota buah dan batang. Bagian nanas yang sering dimanfaatkan adalah bagian daging buah nanas saja sedangkan bagian nanas yang lain tidak dimanfaatkan, ini dikarenakan masih kurangnya pemanfaatan secara maksimal bagian-bagian nanas selain dagingnya. Pada umumnya, bagian tanaman nanas (*ananas comosus*) yang dimanfaatkan hanya buahnya saja, sedangkan bagian lain belum begitu banyak digunakan. Fokus budidaya tanaman nanas adalah untuk diambil buahnya. Selain bisa dimakan secara langsung, buah nanas juga bisa diawetkan melalui pengolahan menjadi beragam produk, seperti jus, selai, dan kripik. Selain itu buah nanas dapat digunakan untuk pelunak daging (Selvia Aprilyanti, 2018: 28-29).

Nanas (*Ananas comosus (L) Merr*) merupakan tanaman buah-buahan yang dapat diproduksi sepanjang tahun, tidak tergantung musim dan termasuk dalam salah satu produk buah unggulan nasional karena hasil produksinya yang melimpah (Wulandari & Daningsih, 2019: 193). Penyebaran tanaman nanas di Indonesia hampir merata terdapat di tiap daerah. Lebih-lebih dalam rangka pembangunan khususnya dibidang pertanian telah digalakkan penye-

baran tanaman nanas, khususnya di daerah transmigrasi di luar pulau Jawa. Melihat tingginya prospek yang menguntungkan, tanaman nanas dapat diharapkan menjadi bahan ekspor non migas, sehingga dapat meningkatkan pendapatan negara di luar produk non migas (Ishak et al., 2017).

2.4.1 Manfaat Nanas

Buah nanas sangat digemari masyarakat karena rasa buahnya yang manis, mengandung cukup vitamin, dan kalori sehingga sangat baik untuk kesehatan. Buah nanas memiliki kandungan enzim bromelin yang bermanfaat melunakkan daging. Ibu yang sedang hamil dilarang makan nanas karena dapat mengakibatkan keguguran.

Enzim bromelin pada buah nanas juga dipercaya sangat baik bagi penderita batuk. Enzim ini dapat menekan rasa gatal pada tenggorokkan yang membuat batuk, mengurangi inflamasi pada hidung, dan memecah lendir. Ini akan membantu melonggarkan tenggorokan. Enzim bromelain juga telah terbukti sangat bermanfaat untuk penyembuhan penyakit infeksi saluran pernapasan seperti bronkitis dan sinusitis. Enzim bromelin pada nanas dapat berperan sebagai anti inflamasi, membantu proses pencernaan dengan melunakkan makanan di lambung, mengganggu pertumbuhan sel kanker, menghambat agregasi platelet, dan mempunyai aktivitas fibrinolitik. Bromelin sering digunakan oleh atlet olahraga untuk mengobati cedera fisik ataupun luka ringan (Putri & Anita, 2017: 490).

2.5 Wahdatul Ulum

Nanas merupakan tanaman yang mempunyai banyak manfaat terutama pada kulitnya, sehingga industri pengolahan buah nanas di Indonesia menjadi prioritas tanaman yang dikembangkan, dan dibuktikan dengan sebagian lahannya dipergunakan sebagai lahan pertanian dan membuat profesi petani mempunyai peran penting. Oleh karena itu Allah menjadikan bumi *dzalul* (mudah dijelajahi) dan *bisath* (hamparan) di mana hal tersebut merupakan nikmat yang harus diingat dan disyukuri. Allah SWT berfirman;

وَأَيُّ لَّهُمُ الْأَرْضُ الْمَيْتَةُ أَحْيَيْهَا وَأَخْرَجْنَا مِنْهَا حَبًّا فَمِنْهُ يَأْكُلُونَ (۳۳)

Artinya: “Dan suatu tanda (kekuasaan Allah yang besar) bagi mereka adalah bumi yang mati. Kami hiduapkan bumi itu dan Kami keluarkan dari padanya biji-bijian, maka daripadanya mereka makan”

وَهُوَ الَّذِي أَنْزَلَ مِنَ السَّمَاءِ مَاءً فَأَخْرَجْنَا بِهِ نَبَاتَ كُلِّ شَيْءٍ فَأَخْرَجْنَا مِنْهُ خَضِرًا نُخْرِجُ مِنْهُ حَبًّا مُتَرَاكِبًا وَمِنَ النَّخْلِ مِنَ طَلْعِهَا قِنْوَانٌ دَانِيَةٌ وَجَنَّاتٍ مِنْ أَعْنَابٍ وَالزَّيْتُونَ وَالزَّيْتُونَ مُشْتَبِهًا وَغَيْرَ مُتَشَبِهٍ أَنْظِرُوا إِلَى ثَمَرِهِ إِذَا أَثْمَرَ وَيَنْعِهِ إِنَّ فِي ذَٰلِكُمْ لَآيَاتٍ لِّقَوْمٍ يُؤْمِنُونَ (٩٩)

Artinya: “Dan Dialah yang menurunkan air hujan dari langit, lalu Kami tumbuhkan dengan air itu segala macam tumbuh-tumbuhan maka Kami keluarkan dari tumbuh-tumbuhan itu tanaman yang menghijau. Kami keluarkan dari tanaman yang menghijau itu butir yang banyak; dan dari mayang korma mengurai tangkai-tangkai yang menjulai, dan kebun-kebun anggur, dan (Kami keluarkan pula) zaitun dan delima yang serupa dan yang tidak serupa. Perhatikanlah buahnya di waktu pohonnya berbuah dan (perhatikan pulalah) kematangannya. Sesungguhnya pada yang demikian itu ada tanda-tanda (kekuasaan Allah) bagi orang-orang yang beriman”

Ayat-ayat di atas merupakan peringatan dari Allah untuk manusia atas nikmat bercocok tanam dan Allah telah memudahkan alat-alat atau keperluannya. Rasulullah Saw pun bersabda:

مَا مِنْ مُسْلِمٍ يَغْرِسُ غَرْسًا أَوْ يَزْرَعُ زَرْعًا فَيَأْكُلُ مِنْهُ طَيْرٌ أَوْ إِنْسَانٌ أَوْ بَيْهِيمَةٌ إِلَّا لَكَانَ لَهُ بِهِ صَدَقَةٌ (رَوَاهُ الْبُخَارِيُّ)

Artinya: “Tidaklah seorang muslim yang menanam tanaman atau bertani kemudian burung, manusia atau pun binatang ternak memakan hasilnya, kecuali semua itu merupakan sedekah baginya”. (HR. Bukhari)

Hadis tersebut menjelaskan bahwa selama hasil tanamannya dimakan oleh burung atau hewan ternak atau dimanfaatkan oleh manusia maka pahala tersebut tetap terus mengalir kepada pemiliknya meskipun ia telah meninggal atau tanamannya berpindah kepemilikan (Rizal, 2021).

Berdasarkan ayat-ayat alquran dan hadits di atas terkandung pesan khusus tentang pertanian dan dapat menjadi pedoman bagi kita untuk meyakini bahwa Allah yang menyediakan irigasi alami berupa air hujan dan karena-Nya kehidupan berlangsung dengan rantai makanan untuk makhluk hidup, sehingga berkembang ilmu pengetahuan dan teknologi.

2.6 Penelitian Terdahulu

Adapun beberapa penelitian sebelumnya yang menggunakan metode *naive bayes* adalah:

1. Berdasarkan penelitian yang telah dilakukan oleh Alfa Saleh pada tahun 2015 berjudul “Implementasi Metode Klasifikasi *Naive Bayes* dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga”.
 - a. Perbedaan tujuan penelitian antara Alfa Saleh dengan penelitian saat ini adalah, penelitian Alfa Saleh bertujuan untuk memprediksi besarnya penggunaan listrik rumah tangga. Sedangkan pada penelitian yang dilakukan peneliti untuk mengetahui nilai tingkat akurasi masalah klasifikasi nanas produktif.
 - b. Hasil penelitian Alfa Saleh yaitu, dari 60 data penggunaan listrik rumah tangga, ada sebanyak 47 data penggunaan listrik rumah tangga berhasil diklasifikasikan dengan benar dan sebanyak 13 data penggunaan listrik rumah tangga tidak berhasil diklasifikasikan dengan benar. Sehingga metode *naive bayes* berhasil memprediksi besarnya penggunaan listrik rumah tangga dengan persentase keakuratan sebesar 78,3333%.
2. Berdasarkan penelitian yang telah dilakukan oleh Haditsah Annur pada tahun 2018 berjudul “Klasifikasi Masyarakat Miskin Menggunakan Metode *Naive Bayes*”.
 - a. Perbedaan tujuan penelitian antara Haditsah Annur dengan penelitian saat ini adalah, penelitian Haditsah Annur bertujuan untuk mengklasifikasi masyarakat miskin di Kecamatan Tibawa menggunakan Algoritma *Naive Bayes*. Sedangkan pada penelitian yang dilakukan peneliti untuk mengetahui nilai tingkat akurasi masalah klasifikasi nanas produktif.

- b. Hasil penelitian Haditsah Annur yaitu, penggunaan metode klasifikasi *naive bayes* terhadap *data set* yang telah diambil pada objek penelitian diperoleh tingkat akurasi sebesar 73% atau termasuk dalam kategori *Good*. Sementara nilai *precision* sebesar 92% dan *recall* sebesar 86%. Berdasarkan hal tersebut dapat dinyatakan bahwa sistem klasifikasi yang dibangun dapat digunakan sebagai bahan masukan bagi pengambil keputusan.
 3. Berdasarkan penelitian yang telah dilakukan oleh Tia Imandasari, Eka Irawan, Agus Perdana Windarto, Anjar Wanto pada tahun 2019 berjudul “Algoritma *Naive Bayes* dalam Klasifikasi Lokasi Pembangunan Sumber Air”.
 - a. Perbedaan tujuan penelitian antara Tia Imandasari, Eka Irawan, Agus Perdana Windarto, Anjar Wanto dengan penelitian saat ini adalah, penelitian Tia Imandasari, dkk bertujuan untuk membantu pihak PDAM Tirta Lihou dalam menentukan lokasi yang layak dilakukan pembangunan sumber air sehingga dapat memenuhi kebutuhan masyarakat. Sedangkan pada penelitian yang dilakukan peneliti untuk mengetahui nilai tingkat akurasi masalah klasifikasi nanas produktif.
 - b. Hasil penelitian yang telah dilakukan Tia Imandasari, Eka Irawan, Agus Perdana Windarto, Anjar Wanto yaitu, total *accuracy* yang diperoleh sebesar 78,95% melalui pengujian data pada *Rapidminer*
- 5.3.
4. Berdasarkan penelitian yang telah dilakukan oleh Ervina Rizka Anandita pada tahun 2014 berjudul “Klasifikasi Nanas Dengan Menggunakan Algoritma *Naive Bayes Classification* Pada Dinas Kehutanan Dan Perkebunan”.
 - a. Perbedaan penelitian Ervina Rizka Anandita dengan penelitian yang dilakukan oleh peneliti adalah, pada penelitian Ervina Rizka dapat meningkatkan level produktivitas perkebunan nanas di Pati, dan meminimalkan biaya produksi untuk membeli varietas nanas yang produktif. Memberikan kemudahan bagi industri nanas untuk meningkatkan pola pembudidayaan, dan produksi nanas agar menghasilkan nanas-nanas yang berkualitas. Sedangkan pada peneli-

tian yang dilakukan peneliti untuk mengetahui nilai tingkat akurasi masalah klasifikasi nanas produktif.

- b. Hasil penelitian Ervina Rizka Anandita yaitu, perhitungan kinerja sistem untuk masalah klasifikasi nanas produktif sebesar 73,3%, dan percobaan yang dilakukan dengan perhitungan probabilitas menghasilkan pengukuran kinerja yang memiliki nilai lebih dari 50%. Maka kinerja *Naive Bayes* yang diterapkan dalam permasalahan ini sudah cukup baik.

Berdasarkan penelitian-penelitian di atas, peneliti dapat menyimpulkan bahwa *data mining* sangat berperan penting dalam pengklasifikasian data yang digunakan untuk menghitung algoritma *naive bayes*. Algoritma *naive bayes* sendiri cukup baik untuk diterapkan dalam penelitian ini.

BAB 3

METODE PENELITIAN

3.1 Tempat dan Waktu Penelitian

Peneliti melaksanakan penelitian di Balai Penyuluhan Pertanian (BPP) Panai Tengah, Kecamatan Panai Tengah, Kabupaten Labuhanbatu. Penentuan lokasi penelitian dilakukan secara sengaja dengan pertimbangan bahwa peningkatan keakuratan pengklasifikasian jenis nanas membutuhkan sebuah perhitungan yang menerapkan metode pengklasifikasian varietas nanas produktif. Waktu penelitian ini dilakukan pada bulan Maret 2021 sampai Februari 2022.

3.2 Jenis Penelitian

Penelitian ini merupakan penelitian deskriptif, metode kuantitatif, dan menggunakan data sekunder yaitu data yang diperoleh melalui Balai Penyuluhan Pertanian (BPP) Panai Tengah. Adapun jenis data yang dibutuhkan meliputi karakter agronomi nanas, yaitu: umur panen (hari), daerah tanam (ha), diameter buah (mm), bobot buah (kg), panjang buah (cm), tinggi tanaman (cm), dan produksi nanas dalam satu periode tanam (kg).

3.3 Variabel Penelitian

Pada penelitian klasifikasi nanas dengan menggunakan algoritma *naive bayes classification* pada balai penyuluhan pertanian (BPP) Panai Tengah menggunakan variabel X dan Y , yakni:

X_1 = Umur panen (hari)

X_2 = Daerah tanam (ha)

X_3 = Diameter buah (mm)

X_4 = Bobot buah (kg)

X_5 = Panjang buah (cm)

X_6 = Tinggi tanaman (cm)

X_7 = Jumlah produksi nanas dalam satu periode tanam (kg)

Y = Keterangan

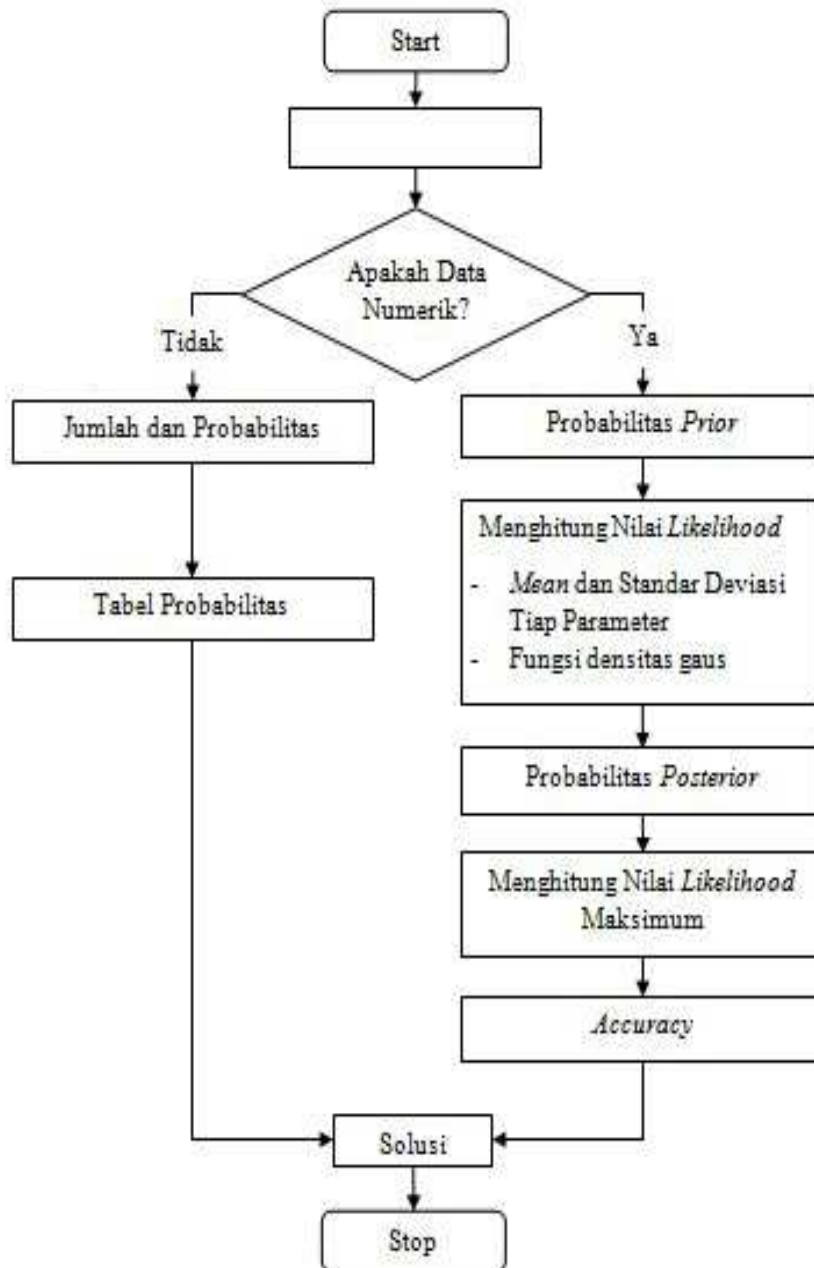
3.4 Prosedur Penelitian

Adapun tahapan yang dilakukan pada rancang bangun sistem pendukung penentuan klasifikasi nanas sebagai berikut:

1. Baca *data training*, namun apabila data numerik maka:
2. Hitung probabilitas *prior*.
3. Mencari probabilitas bersyarat (*likelihood*), *likelihood* dapat dihitung dengan mengalikan hasil dari masing-masing nilai probabilitas per-atribut yang berdasarkan pada w_j . Langkah-langkah menghitung *likelihood*:
 - a. Cari nilai mean dan standar deviasi dari masing-masing parameter yang merupakan data numerik, dan
 - b. Persamaan untuk menghitung fungsi densitas gauss
4. Hitung probabilitas *posterior*
5. Mendapatkan nilai maksimum *likelihood*.
6. Nilai *accuracy*
7. Solusi kemudian dihasilkan.

3.5 Diagram Alir

Berikut adalah alur metode *naive bayes* yang akan diterapkan:



Gambar 3.1 Diagram Alir

BAB 4

HASIL DAN PEMBAHASAN

Berikut merupakan penjelasan mengenai hasil penelitian berdasarkan data klasifikasi nanas menggunakan *naive bayes classification* pada Balai Penyuluhan Pertanian (BPP) Panai Tengah.

4.1 Pengumpulan Data

Pada tahap ini diawali dengan melakukan pengambilan data sampel dari instansi Balai Penyuluhan Pertanian (BPP) Panai Tengah, nantinya akan dijadikan sebagai *data training*. Pengujian menggunakan data nanas di Kecamatan Panai Tengah, data nanas yang diperoleh sebanyak 86 *record* dan diambil 52 *record* untuk digunakan sebagai *data training*. Berdasarkan hasil pengolahan data yang diperoleh dari Balai Penyuluhan Pertanian (BPP) Panai Tengah dibagi menjadi dua kategori yaitu: Ya sebanyak 31 nanas dan Tidak sebanyak 21 nanas.

Proses pengujian, data dibagi menjadi 2 bagian yaitu *data training* dan *data testing* dengan menggunakan *naive bayes*. *Data training* digunakan untuk membentuk tabel probabilitas dan *data testing* digunakan untuk menguji probabilitas yang telah terbentuk. Data tersebut dapat dilihat pada Tabel 4.1 di bawah ini. Selanjutnya menganalisis metode *naive bayes classification*:

Tabel 4.1 : *Data Training*

No	Jenis Nanas	X_1	X_2	X_3	X_4	X_5	X_6	X_7	Y
1	An1	252	8	90,3	1,35	17	99	6480	Ya
2	An3	252	8	99,1	1,4	20	91	5400	Ya
3	An4	280	10	96,1	1,55	22	74	8100	Ya
4	An7	238	8	87,9	1,37	18	97	5400	Tidak
5	An12	238	8	84,9	0,98	15	109	8600	Tidak
6	An13	238	8	86,6	1,25	17	96	6500	Ya
7	An14	392	5	92,9	0,83	18	128	2700	Tidak
8	An16	336	8	79,2	0,74	10	73	4400	Tidak
9	An17	336	8	81,7	0,86	14	97	6500	Tidak
10	An18	392	10	93,1	1,16	17,5	81	5400	Ya
11	An19	280	10	87,8	0,9	13,5	96	5400	Tidak
12	An20	280	12	93,3	1	16	100	6400	Ya
13	An22	252	12	91	1,5	21	125	6400	Ya
14	An23	280	6	80,6	0,69	14	62	3000	Tidak
15	An25	336	8	95,8	1,2	16	128	4300	Ya
16	An26	336	12	94,3	1,14	17	94	6400	Ya
17	An28	238	10	96,1	1,4	19	90	5400	Ya
18	An31	280	8	82,2	0,68	12	99	4300	Tidak
19	An32	392	10	93	1	16	99	5400	Ya
20	An33	392	10	89,9	0,98	13	67	5400	Tidak
21	An35	252	6	94,2	0,86	14	110	3100	Tidak
22	An36	252	10	92,5	0,95	14	101	5400	Tidak
23	An37	336	10	92,8	1,37	17,5	91	5400	Tidak
24	An38	336	8	92,6	1,2	15	90	4400	Ya
25	An39	336	8	95,2	1,15	15	67	4000	Ya
26	An40	336	8	98,9	1,15	15,5	82	4320	Ya
27	An42	392	4	93,8	1,14	16,5	71	3000	Ya
28	An45	392	6	16,5	1,25	16,5	67	3200	Ya
29	An49	252	10	95,5	1,23	16,5	99	5400	Ya
30	An50	280	10	89,2	1,14	16,5	99	5400	Ya
31	An51	252	12	92,6	1	15,5	82	6400	Tidak
32	An52	280	12	87,9	0,9	18	67	6400	Tidak
33	An53	280	8	87,6	0,94	13,5	62	7500	Tidak
34	An55	280	4	85,2	0,7	14	62	3000	Tidak
35	An56	336	5	91,8	1,24	18	99	2700	Ya
36	An58	336	4	82	0,75	13,5	110	2160	Tidak
37	An59	252	10	93,3	1,16	16	99	5400	Ya
38	An61	252	12	87,4	0,7	13,5	87	6480	Tidak
39	An65	392	5	85	0,73	13	93	2700	Tidak
40	An66	252	5	92	1,35	21	99	2700	Ya
41	An68	336	4	99,7	1,25	15,5	67	2500	Ya
42	An71	252	8	92,2	1,26	17	67	7600	Ya
43	An73	392	12	95,5	1,1	15	65	5400	Ya
44	An74	252	6	92,9	0,92	16	67	3400	Tidak
45	An76	336	4	93,8	1,8	15,8	112	2100	Ya
46	An78	336	6	96,1	1,1	15	110	8400	Ya
47	An79	336	8	92,4	1,15	16,5	67	8500	Ya
48	An82	280	8	99,4	1,19	16,8	87	4400	Ya
49	An83	392	5	91,8	1,39	18,5	67	2700	Ya
50	An84	392	6	97,8	1,39	19	86	6600	Ya
51	An85	280	5	97	1,1	15	101	2700	Tidak
52	An86	252	4	92,6	1,23	17	99	2160	Ya

4.2 Baca *Data Training* Tabel 4.1

Langkah pertama yang dilakukan adalah membaca *data training* kemudian mengelompokkan variabel berdasarkan klasifikasi nanas antara data diskrit dan data kontinu. Data yang diperoleh diketahui bahwa tidak terdapat data diskrit dan hanya terdapat tujuh data kontinu, diantaranya:

- a. Data kontinu
 - Umur panen (hari)
 - Daerah tanam (ha)
 - Diameter buah (mm)
 - Bobot buah (kg)
 - Panjang buah (cm)
 - Produksi nanas dalam satu periode tanam (kg).

4.3 Menghitung Probabilitas *Prior*

Kemudian langkah kedua pencarian klasifikasi dengan metode *naive bayes*, yaitu: mencari probabilitas dari masing-masing nanas. Prediksi keproduktifan nanas akan ditentukan oleh dua kategori yaitu “Ya” dan “Tidak”. Perhitungan probabilitas dengan cara mencari jumlah data Ya dan Tidak dari total keseluruhan *data training*, lalu membaginya dengan keseluruhan data.

Tabel 4.2 : Probabilitas *Prior*

Probabilitas Nanas			
		Probabilitas	
YA	TIDAK	YA	TIDAK
31	21	31/52	21/52

4.4 Menghitung Nilai *Likelihood*

Likelihood dapat dihitung dengan mengalikan hasil dari masing-masing nilai probabilitas setiap atribut yang berdasarkan pada w_j . Langkah-langkah menghitung *Likelihood*:

4.4.1 Menghitung Nilai *Mean* dan Standar Deviasi

Langkah selanjutnya yang harus dilakukan adalah menentukan nilai rata-rata atau *mean* (μ) dan standar deviasi (σ) setiap atribut, diantaranya: umur panen (hari), daerah tanam (ha), diameter buah (mm), bobot buah (kg), panjang buah (cm), dan produksi nanas dalam satu periode tanam (kg).

Mean dan standar deviasi untuk masing-masing kategori (Ya dan Tidak) dari ketujuh atribut tersebut sebagai berikut:

1. Umur Panen (Hari)

$$\begin{aligned}\mu_{ya} &= \frac{252 + 252 + 280 + 238 + 392 + 280 + 252 + 336 + 336 + 238 + \\ & 392 + 336 + 336 + 336 + 392 + 392 + 252 + 280 + 336 + 252 + \\ & 252 + 336 + 252 + 392 + 336 + 336 + 336 + 280 + 392 + 392 + \\ & 252}{31} \\ &= \frac{9716}{31} \\ &= 313,4193548\end{aligned}$$

$$\begin{aligned}\mu_{tidak} &= \frac{238 + 238 + 392 + 336 + 336 + 280 + 280 + 280 + 392 + 252 + \\ & 252 + 336 + 252 + 280 + 280 + 280 + 336 + 252 + 392 + 252 + \\ & 280}{21} \\ &= \frac{6216}{21} \\ &= 296\end{aligned}$$

$$\begin{aligned}\sigma_{ya}^2 &= \frac{(252 - 313,4193548)^2 + (252 - 313,4193548)^2 + \\ & (280 - 313,4193548)^2 + (238 - 313,4193548)^2 + \\ & (392 - 313,4193548)^2 + (280 - 313,4193548)^2 + \\ & (252 - 313,4193548)^2 + (336 - 313,4193548)^2 + \\ & (336 - 313,4193548)^2 + (238 - 313,4193548)^2 + \\ & (392 - 313,4193548)^2 + (336 - 313,4193548)^2 + \\ & (336 - 313,4193548)^2 + (336 - 313,4193548)^2 + \\ & (392 - 313,4193548)^2 + (392 - 313,4193548)^2 + \\ & (252 - 313,4193548)^2 + (280 - 313,4193548)^2 + \\ & (336 - 313,4193548)^2 + (252 - 313,4193548)^2 + \\ & (252 - 313,4193548)^2 + (336 - 313,4193548)^2 + \\ & (252 - 313,4193548)^2 + (392 - 313,4193548)^2 + \\ & (336 - 313,4193548)^2 + (336 - 313,4193548)^2 + \\ & (336 - 313,4193548)^2 + (280 - 313,4193548)^2 + \\ & (392 - 313,4193548)^2 + (392 - 313,4193548)^2 + \\ & (252 - 313,4193548)^2}{31-1}\end{aligned}$$

$$\begin{aligned}
& 3772,337149 + 3772,337149 + 1116,853278 + 5688,079084 + \\
& 6174,917794 + 1116,853278 + 3772,337149 + 509,8855359 + \\
& 509,8855359 + 5688,079084 + 6174,917794 + 509,8855359 + \\
& 509,8855359 + 509,8855359 + 6174,917794 + 6174,917794 + \\
& 3772,337149 + 1116,853278 + 509,8855359 + 3772,337149 + \\
& 3772,337149 + 509,8855359 + 3772,337149 + 6174,917794 + \\
& 509,8855359 + 509,8855359 + 509,8855359 + 1116,853278 + \\
& 6174,917794 + 6174,917794 + 3772,337149 \\
& = \frac{\quad}{30} \\
& = \frac{94345,54839}{30} \\
& = 3144,851613 \\
\sigma_{ya} &= \sqrt{3144,851613} \\
& = 56,07897657 \\
& (238 - 296)^2 + (238 - 296)^2 + (392 - 296)^2 + (336 - 296)^2 + \\
& (336 - 296)^2 + (280 - 296)^2 + (280 - 296)^2 + (280 - 296)^2 + \\
& (392 - 296)^2 + (252 - 296)^2 + (252 - 296)^2 + (336 - 296)^2 + \\
& (252 - 296)^2 + (280 - 296)^2 + (280 - 296)^2 + (280 - 296)^2 + \\
& (336 - 296)^2 + (252 - 296)^2 + (392 - 296)^2 + (252 - 296)^2 + \\
& (280 - 296)^2 \\
\sigma_{tidak}^2 &= \frac{\quad}{21-1} \\
& \frac{3364 + 3364 + 9216 + 1600 + 1600 + 256 + 256 + 256 + 9216 + \\
& 1936 + 1936 + 1600 + 1936 + 256 + 256 + 256 + 1600 + 1936 + \\
& 9216 + 1936 + 256}{20} \\
& = \frac{52248}{20} \\
& = 2612,4 \\
\sigma_{tidak} &= \sqrt{2612,4} \\
& = 51,11164251
\end{aligned}$$

Dari hasil perhitungan umur panen diperoleh hasil yang menunjukkan nilai *mean* Ya = 313,4193548, standar deviasi Ya = 56,07897657, dan hasil *mean* Tidak = 296, standar deviasi Tidak = 51,11164251.

2. Daerah Tanam (Ha)

$$\begin{aligned}
\mu_{ya} &= \frac{8 + 8 + 10 + 8 + 10 + 12 + 12 + 8 + 12 + 10 + 10 + 8 + 8 + 8 + 4 + \\
& 6 + 10 + 10 + 5 + 10 + 5 + 4 + 8 + 12 + 4 + 6 + 8 + 8 + 5 + 6 + 4}{31} \\
& = \frac{247}{31} \\
& = 7,967741935
\end{aligned}$$

$$\mu_{tidak} = \frac{8 + 8 + 5 + 8 + 8 + 10 + 6 + 8 + 10 + 6 + 10 + 10 + 12 + 12 + 8 + 4 + 4 + 12 + 5 + 6 + 5}{21}$$

$$= \frac{165}{21}$$

$$= 7,857142857$$

$$\sigma_{ya}^2 = \frac{(8 - 7,967741935)^2 + (8 - 7,967741935)^2 + (10 - 7,967741935)^2 + (8 - 7,967741935)^2 + (10 - 7,967741935)^2 + (12 - 7,967741935)^2 + (12 - 7,967741935)^2 + (8 - 7,967741935)^2 + (12 - 7,967741935)^2 + (10 - 7,967741935)^2 + (10 - 7,967741935)^2 + (8 - 7,967741935)^2 + (8 - 7,967741935)^2 + (4 - 7,967741935)^2 + (6 - 7,967741935)^2 + (10 - 7,967741935)^2 + (10 - 7,967741935)^2 + (5 - 7,967741935)^2 + (10 - 7,967741935)^2 + (5 - 7,967741935)^2 + (4 - 7,967741935)^2 + (8 - 7,967741935)^2 + (12 - 7,967741935)^2 + (4 - 7,967741935)^2 + (6 - 7,967741935)^2 + (8 - 7,967741935)^2 + (8 - 7,967741935)^2 + (5 - 7,967741935)^2 + (6 - 7,967741935)^2 + (4 - 7,967741935)^2}{31-1}$$

$$= \frac{0,001040583 + 0,001040583 + 4,130072841 + 0,001040583 + 4,130072841 + 16,2591051 + 16,2591051 + 0,001040583 + 16,2591051 + 4,130072841 + 4,130072841 + 0,001040583 + 0,001040583 + 0,001040583 + 15,74297607 + 3,872008325 + 4,130072841 + 4,130072841 + 8,807492196 + 4,130072841 + 8,807492196 + 15,74297607 + 0,001040583 + 16,2591051 + 15,74297607 + 3,872008325 + 0,001040583 + 0,001040583 + 8,807492196 + 3,872008325 + 15,74297607}{30}$$

$$= \frac{194,9677419}{30}$$

$$= 6,498924731$$

$$\sigma_{ya} = \sqrt{6,498924731}$$

$$= 2,549298871$$

$$\sigma_{tidak}^2 = \frac{(8 - 7,857142857)^2 + (8 - 7,857142857)^2 + (5 - 7,857142857)^2 + (8 - 7,857142857)^2 + (8 - 7,857142857)^2 + (10 - 7,857142857)^2 + (6 - 7,857142857)^2 + (8 - 7,857142857)^2 + (10 - 7,857142857)^2 + (6 - 7,857142857)^2 + (10 - 7,857142857)^2 + (10 - 7,857142857)^2 + (12 - 7,857142857)^2 + (12 - 7,857142857)^2 + (8 - 7,857142857)^2 + (4 - 7,857142857)^2 + (4 - 7,857142857)^2 + (12 - 7,857142857)^2 + (5 - 7,857142857)^2 + (6 - 7,857142857)^2 + (5 - 7,857142857)^2}{21-1}$$

$$= \frac{134,5714286}{20}$$

$$= 6,728571429$$

$$\sigma_{tidak} = \sqrt{6,728571429}$$

$$= 2,593949003$$

Dari hasil perhitungan daerah tanam diperoleh hasil yang menunjukkan nilai *mean* Ya = 7,967741935, standar deviasi Ya = 2,549298871, dan hasil *mean* Tidak = 7,857142857, standar deviasi Tidak = 2,593949003.

3. Diameter Buah (Mm)

$$\begin{aligned}\mu_{ya} &= \frac{90,3 + 99,1 + 96,1 + 86,6 + 93,1 + 93,3 + 91 + 95,8 + 94,3 + \\ &96,1 + 93 + 92,6 + 95,2 + 98,9 + 93,8 + 16,5 + 95,5 + 89,2 + \\ &91,8 + 93,3 + 92 + 99,7 + 92,2 + 95,5 + 93,8 + 96,1 + 92,4 + \\ &99,4 + 91,8 + 97,8 + 92,6}{31} \\ &= \frac{2838,8}{31} \\ &= 91,57419355\end{aligned}$$

$$\begin{aligned}\mu_{tidak} &= \frac{87,9 + 84,9 + 92,9 + 79,2 + 81,7 + 87,8 + 80,6 + 82,2 + \\ &89,9 + 94,2 + 92,5 + 92,8 + 92,6 + 87,9 + 87,6 + 85,2 + \\ &82 + 87,4 + 85 + 92,9 + 97}{21} \\ &= \frac{1844,2}{21} \\ &= 87,81904762\end{aligned}$$

$$\begin{aligned}\sigma_{ya}^2 &= \frac{(90,3 - 91,57419355)^2 + (99,1 - 91,57419355)^2 + \\ &(96,1 - 91,57419355)^2 + (86,6 - 91,57419355)^2 + \\ &(93,1 - 91,57419355)^2 + (93,3 - 91,57419355)^2 + \\ &(91 - 91,57419355)^2 + (95,8 - 91,57419355)^2 + \\ &(94,3 - 91,57419355)^2 + (96,1 - 91,57419355)^2 + \\ &(93 - 91,57419355)^2 + (92,6 - 91,57419355)^2 + \\ &(95,2 - 91,57419355)^2 + (98,9 - 91,57419355)^2 + \\ &(93,8 - 91,57419355)^2 + (16,5 - 91,57419355)^2 + \\ &(95,5 - 91,57419355)^2 + (89,2 - 91,57419355)^2 + \\ &(91,8 - 91,57419355)^2 + (93,3 - 91,57419355)^2 + \\ &(92 - 91,57419355)^2 + (99,7 - 91,57419355)^2 + \\ &(92,2 - 91,57419355)^2 + (95,5 - 91,57419355)^2 + \\ &(93,8 - 91,57419355)^2 + (96,1 - 91,57419355)^2 + \\ &(92,4 - 91,57419355)^2 + (99,4 - 91,57419355)^2 + \\ &(91,8 - 91,57419355)^2 + (97,8 - 91,57419355)^2 + \\ &(92,6 - 91,57419355)^2}{31-1} \\ &= \frac{1,623569199 + 56,63776275 + 20,48292404 + 24,74260146 + \\ &2,328085328 + 2,978407908 + 0,329698231 + 17,85744017 + \\ &7,430020812 + 20,48292404 + 2,032924037 + 1,052278876 + \\ &13,14647242 + 53,66744017 + 4,95421436 + 5636,134537 + \\ &15,4119563 + 5,636795005 + 0,050988554 + 2,978407908 + \\ &0,181311134 + 66,02873049 + 0,391633715 + 15,4119563 + \\ &4,95421436 + 20,48292404 + 0,681956296 + 61,24324662 + \\ &0,050988554 + 38,76066597 + 1,052278876}{30} \\ &= \frac{6099,199355}{30} \\ &= 203,3066452\end{aligned}$$

$$\begin{aligned}
&= \frac{0,766580952}{20} \\
&= 0,038329048 \\
\sigma_{tidak} &= \sqrt{0,038329048} \\
&= 0,195778057
\end{aligned}$$

Dari hasil perhitungan bobot buah diperoleh hasil yang menunjukkan nilai *mean* Ya = 1,252258065, standar deviasi Ya = 0,165443035, dan hasil *mean* Tidak = 0,902380952, standar deviasi Tidak = 0,195778057.

5. Panjang Buah (Cm)

$$\begin{aligned}
\mu_{ya} &= \frac{17 + 20 + 22 + 17 + 17,5 + 16 + 21 + 16 + 17 + 19 + 16 + 15 + 15 + 15,5 + 16,5 + 16,5 + 16,5 + 16,5 + 18 + 16 + 21 + 15,5 + 17 + 15 + 15,8 + 15 + 16,5 + 16,8 + 18,5 + 19 + 17}{31} \\
&= \frac{531,1}{31} \\
&= 17,13225806
\end{aligned}$$

$$\begin{aligned}
\mu_{tidak} &= \frac{18 + 15 + 18 + 10 + 14 + 13,5 + 14 + 12 + 13 + 14 + 14 + 17,5 + 15,5 + 18 + 13,5 + 14 + 13,5 + 13,5 + 13 + 16 + 15}{21} \\
&= \frac{305}{21} \\
&= 14,52380952
\end{aligned}$$

$$\begin{aligned}
\sigma_{ya}^2 &= \frac{(17 - 17,13225806)^2 + (20 - 17,13225806)^2 + (22 - 17,13225806)^2 + (17 - 17,13225806)^2 + (17,5 - 17,13225806)^2 + (16 - 17,13225806)^2 + (21 - 17,13225806)^2 + (16 - 17,13225806)^2 + (17 - 17,13225806)^2 + (19 - 17,13225806)^2 + (16 - 17,13225806)^2 + (15 - 17,13225806)^2 + (15,5 - 17,13225806)^2 + (16,5 - 17,13225806)^2 + (16,5 - 17,13225806)^2 + (16,5 - 17,13225806)^2 + (18 - 17,13225806)^2 + (16 - 17,13225806)^2 + (21 - 17,13225806)^2 + (15,5 - 17,13225806)^2 + (17 - 17,13225806)^2 + (15 - 17,13225806)^2 + (15,8 - 17,13225806)^2 + (15 - 17,13225806)^2 + (16,5 - 17,13225806)^2 + (16,8 - 17,13225806)^2 + (18,5 - 17,13225806)^2 + (19 - 17,13225806)^2 + (17 - 17,13225806)^2}{31-1} \\
&= \frac{0,017492196 + 8,223943809 + 23,69491155 + 0,017492196 + 0,135234131 + 1,282008325 + 14,95942768 + 1,282008325 + 0,017492196 + 3,488459938 + 1,282008325 + 4,546524454 + 4,546524454 + 2,664266389 + 0,39975026 + 0,39975026 + 0,39975026 + 0,39975026 + 0,752976067 + 1,282008325 + 14,95942768 + 2,664266389 + 0,017492196 + 4,546524454 + 1,77491155 + 4,546524454 + 0,39975026 + 0,110395421 + 1,870718002 + 3,488459938 + 0,017492196}{30}
\end{aligned}$$

$$\begin{aligned}
&= \frac{104,1877419}{30} \\
&= 3,472924731 \\
\sigma_{ya} &= \sqrt{3,472924731} \\
&= 1,863578475 \\
&\quad (18 - 14,52380952)^2 + (15 - 14,52380952)^2 + \\
&\quad (18 - 14,52380952)^2 + (10 - 14,52380952)^2 + \\
&\quad (14 - 14,52380952)^2 + (13,5 - 14,52380952)^2 + \\
&\quad (14 - 14,52380952)^2 + (12 - 14,52380952)^2 + \\
&\quad (13 - 14,52380952)^2 + (14 - 14,52380952)^2 + \\
&\quad (14 - 14,52380952)^2 + (17,5 - 14,52380952)^2 + \\
&\quad (15,5 - 14,52380952)^2 + (18 - 14,52380952)^2 + \\
&\quad (13,5 - 14,52380952)^2 + (14 - 14,52380952)^2 + \\
&\quad (13,5 - 14,52380952)^2 + (13,5 - 14,52380952)^2 + \\
&\quad (13 - 14,52380952)^2 + (16 - 14,52380952)^2 + \\
&\quad (15 - 14,52380952)^2 \\
\sigma_{tidak}^2 &= \frac{ \\
&\quad 12,08390023 + 0,22675737 + 12,08390023 + 20,46485261 + \\
&\quad 0,274376417 + 1,048185941 + 0,274376417 + 6,369614512 + \\
&\quad 2,321995465 + 0,274376417 + 0,274376417 + 8,857709751 + \\
&\quad 0,952947846 + 12,08390023 + 1,048185941 + 0,274376417 + \\
&\quad 1,048185941 + 1,048185941 + 2,321995465 + 2,179138322 + \\
&\quad 0,22675737}{21-1} \\
&= \frac{ \\
&\quad 12,08390023 + 0,22675737 + 12,08390023 + 20,46485261 + \\
&\quad 0,274376417 + 1,048185941 + 0,274376417 + 6,369614512 + \\
&\quad 2,321995465 + 0,274376417 + 0,274376417 + 8,857709751 + \\
&\quad 0,952947846 + 12,08390023 + 1,048185941 + 0,274376417 + \\
&\quad 1,048185941 + 1,048185941 + 2,321995465 + 2,179138322 + \\
&\quad 0,22675737}{20} \\
&= \frac{85,73809524}{20} \\
&= 4,286904762 \\
\sigma_{tidak} &= \sqrt{4,286904762} \\
&= 2,070484185
\end{aligned}$$

Dari hasil perhitungan panjang buah diperoleh hasil yang menunjukkan nilai *mean* Ya = 17,13225806, standar deviasi Ya = 1,863578475, dan hasil *mean* Tidak = 14,52380952, standar deviasi Tidak = 2,070484185.

6. Tinggi Tanaman (Cm)

$$\begin{aligned}
\mu_{ya} &= \frac{99 + 91 + 74 + 96 + 81 + 100 + 125 + 128 + 94 + 90 + 99 + 90 + \\
&\quad 67 + 82 + 71 + 67 + 99 + 99 + 99 + 99 + 99 + 67 + 67 + 65 + \\
&\quad 112 + 110 + 67 + 87 + 67 + 86 + 99}{31} \\
&= \frac{2776}{31} \\
&= 89,5483871
\end{aligned}$$

$$\begin{aligned}
\mu_{tidak} &= \frac{97 + 109 + 128 + 73 + 97 + 96 + 62 + 99 + 110 + 101 + 91 + 82 + 67 + 62 + 62 + 110 + 87 + 93 + 67 + 101}{21} \\
&= \frac{1861}{21} \\
&= 88,61904762 \\
\sigma_{ya}^2 &= \frac{(99 - 88,61904762)^2 + (91 - 88,61904762)^2 + (74 - 88,61904762)^2 + (96 - 88,61904762)^2 + (81 - 88,61904762)^2 + (100 - 88,61904762)^2 + (125 - 88,61904762)^2 + (128 - 88,61904762)^2 + (94 - 88,61904762)^2 + (90 - 88,61904762)^2 + (99 - 88,61904762)^2 + (90 - 88,61904762)^2 + (67 - 88,61904762)^2 + (82 - 88,61904762)^2 + (71 - 88,61904762)^2 + (67 - 88,61904762)^2 + (99 - 88,61904762)^2 + (99 - 88,61904762)^2 + (99 - 88,61904762)^2 + (67 - 88,61904762)^2 + (67 - 88,61904762)^2 + (65 - 88,61904762)^2 + (112 - 88,61904762)^2 + (110 - 88,61904762)^2 + (67 - 88,61904762)^2 + (87 - 88,61904762)^2 + (67 - 88,61904762)^2 + (86 - 88,61904762)^2 + (99 - 88,61904762)^2}{31-1} \\
&= \frac{89,33298647 + 2,107180021 + 241,7523413 + 41,62330905 + 73,07492196 + 109,2362123 + 1256,816857 + 1478,526535 + 19,81685744 + 0,203954214 + 89,33298647 + 0,203954214 + 508,4297607 + 56,97814776 + 344,0426639 + 508,4297607 + 89,33298647 + 89,33298647 + 89,33298647 + 89,33298647 + 89,33298647 + 508,4297607 + 508,4297607 + 602,6233091 + 504,074922 + 418,2684703 + 508,4297607 + 6,494276795 + 508,4297607 + 12,59105099 + 89,33298647}{30} \\
&= \frac{8933,677419}{30} \\
&= 297,7892473 \\
\sigma_{ya} &= \sqrt{297,7892473} \\
&= 17,25657113 \\
\sigma_{tidak}^2 &= \frac{(97 - 88,61904762)^2 + (109 - 88,61904762)^2 + (128 - 88,61904762)^2 + (73 - 88,61904762)^2 + (97 - 88,61904762)^2 + (96 - 88,61904762)^2 + (62 - 88,61904762)^2 + (99 - 88,61904762)^2 + (67 - 88,61904762)^2 + (110 - 88,61904762)^2 + (101 - 88,61904762)^2 + (91 - 88,61904762)^2 + (82 - 88,61904762)^2 + (67 - 88,61904762)^2 + (62 - 88,61904762)^2 + (62 - 88,61904762)^2 + (110 - 88,61904762)^2 + (87 - 88,61904762)^2 + (93 - 88,61904762)^2 + (67 - 88,61904762)^2 + (101 - 88,61904762)^2}{21-1} \\
&= \frac{70,24036281 + 415,38322 + 1550,85941 + 243,9546485 + 70,24036281 + 54,47845805 + 708,5736961 + 107,7641723 + 467,38322 + 457,1451247 + 153,2879819 + 5,66893424 + 43,81179138 + 467,38322 + 708,5736961 + 708,5736961 + 457,1451247 + 2,621315193 + 19,19274376 + 467,38322 + 153,2879819}{20}
\end{aligned}$$

$$\begin{aligned}
&= \frac{7332,952381}{20} \\
&= 366,647619 \\
\sigma_{tidak} &= \sqrt{366,647619} \\
&= 19,14804478
\end{aligned}$$

Dari hasil perhitungan tinggi tanaman diperoleh hasil yang menunjukkan nilai *mean* Ya = 89,5483871, standar deviasi Ya = 17,25657113, dan hasil *mean* Tidak = 88,61904762, standar deviasi Tidak = 19,14804478.

7. Produksi Nanas (Kg)

$$\begin{aligned}
\mu_{ya} &= \frac{6480 + 5400 + 8100 + 6500 + 5400 + 6400 + 6400 + 4300 + 6400 + 5400 + 5400 + 4400 + 4000 + 4320 + 3000 + 3200 + 5400 + 5400 + 2700 + 5400 + 2700 + 2500 + 7600 + 5400 + 2100 + 8400 + 8500 + 4400 + 2700 + 6600 + 2160}{31} \\
&= \frac{157060}{31} \\
&= 5066,451613 \\
\mu_{tidak} &= \frac{5400 + 8600 + 2700 + 4400 + 6500 + 5400 + 3000 + 4300 + 5400 + 3100 + 5400 + 5400 + 6400 + 6400 + 7500 + 3000 + 2160 + 6480 + 2700 + 3400 + 2700}{21} \\
&= \frac{100340}{21} \\
&= 4778,095238 \\
\sigma_{ya}^2 &= \frac{(6480 - 5066,451613)^2 + (5400 - 5066,451613)^2 + (8100 - 5066,451613)^2 + (6500 - 5066,451613)^2 + (5400 - 5066,451613)^2 + (6400 - 5066,451613)^2 + (6400 - 5066,451613)^2 + (4300 - 5066,451613)^2 + (6400 - 5066,451613)^2 + (5400 - 5066,451613)^2 + (5400 - 5066,451613)^2 + (4400 - 5066,451613)^2 + (4000 - 5066,451613)^2 + (4320 - 5066,451613)^2 + (3000 - 5066,451613)^2 + (3200 - 5066,451613)^2 + (5400 - 5066,451613)^2 + (5400 - 5066,451613)^2 + (2700 - 5066,451613)^2 + (5400 - 5066,451613)^2 + (2700 - 5066,451613)^2 + (2500 - 5066,451613)^2 + (7600 - 5066,451613)^2 + (5400 - 5066,451613)^2 + (2100 - 5066,451613)^2 + (8400 - 5066,451613)^2 + (8500 - 5066,451613)^2 + (4400 - 5066,451613)^2 + (2700 - 5066,451613)^2 + (6600 - 5066,451613)^2 + (2160 - 5066,451613)^2}{31-1}
\end{aligned}$$

$$\begin{aligned}
& 1998119,043 + 111254,5265 + 9202415,817 + \\
& 2055060,978 + 111254,5265 + 1778351,301 + \\
& 1778351,301 + 587448,0749 + 1778351,301 + \\
& 111254,5265 + 111254,5265 + 444157,7523 + \\
& 1137319,043 + 557190,0104 + 4270222,268 + \\
& 3483641,623 + 111254,5265 + 111254,5265 + \\
& 5600093,236 + 111254,5265 + 5600093,236 + \\
& 6586673,881 + 6418867,43 + 111254,5265 + \\
& 8799835,172 + 11112544,85 + 11789254,53 + \\
& 444157,7523 + 5600093,236 + 2351770,656 + \\
& 8447460,978 \\
= & \frac{\hspace{15em}}{30} \\
= & \frac{102711509,7}{30} \\
= & 3423716,989 \\
\sigma_{ya} = & \sqrt{3423716,989} \\
= & 1850,328887 \\
& (5400 - 4778,095238)^2 + (8600 - 4778,095238)^2 + \\
& (2700 - 4778,095238)^2 + (4400 - 4778,095238)^2 + \\
& (6500 - 4778,095238)^2 + (5400 - 4778,095238)^2 + \\
& (3000 - 4778,095238)^2 + (4300 - 4778,095238)^2 + \\
& (5400 - 4778,095238)^2 + (3100 - 4778,095238)^2 + \\
& (5400 - 4778,095238)^2 + (5400 - 4778,095238)^2 + \\
& (6400 - 4778,095238)^2 + (6400 - 4778,095238)^2 + \\
& (7500 - 4778,095238)^2 + (3000 - 4778,095238)^2 + \\
& (2160 - 4778,095238)^2 + (6480 - 4778,095238)^2 + \\
& (2700 - 4778,095238)^2 + (3400 - 4778,095238)^2 + \\
& (2700 - 4778,095238)^2 \\
\sigma_{tidak}^2 = & \frac{\hspace{15em}}{21-1} \\
& 386765,5329 + 14606956,01 + 4318479,819 + \\
& 142956,0091 + 2964956,009 + 386765,5329 + \\
& 3161622,676 + 228575,0567 + 386765,5329 + \\
& 2816003,628 + 386765,5329 + 386765,5329 + \\
& 2630575,057 + 2630575,057 + 7408765,533 + \\
& 3161622,676 + 6854422,676 + 2896479,819 + \\
& 4318479,819 + 1899146,485 + 4318479,819 \\
= & \frac{\hspace{15em}}{20} \\
= & \frac{66291923,81}{20} \\
= & 3314596,19 \\
\sigma_{tidak} = & \sqrt{3314596,19} \\
= & 1820,603249
\end{aligned}$$

Dari hasil perhitungan produksi nanas diperoleh hasil yang menunjukkan nilai *mean* Ya = 5066,451613, standar deviasi Ya = 1850,328887, dan hasil *mean* Tidak = 4778,095238, standar deviasi Tidak = 1820,603249.

4.4.2 Perhitungan prediksi dengan *Naive Bayes* dengan Fungsi Densitas Gauss

Setelah diperoleh nilai *mean* dan standar deviasi setiap atribut, maka tahap selanjutnya adalah: menghitung metode *Naive Bayes* dengan rumus dentitas Gauss. Untuk mengklasifikasi nanas produktif atau tidak produktif, sebagai contoh jika diketahui: jenis nanas, umur panen 448 hari, daerah tanam 12 ha, diameter buah 95 mm, bobot buah 1,8 kg, panjang buah 25 cm, tinggi tanman 97 cm, produksi nanas 9000 buah.

1. Umur Panen (Hari)

$$\begin{aligned}
 P(x_1 = |Ya) &= \frac{1}{\sqrt{2.3,14.56,07897657}} 2,718282^{\frac{-(448-313,4193548)^2}{2(56,07897657)^2}} \\
 &= \frac{1}{\sqrt{352,1759728}} 2,718282^{\frac{-(134,5806452)^2}{2(3144,851613)}} \\
 &= \frac{1}{18,76635215} 2,718282^{\frac{-18111,95005}{6289,703226}} \\
 &= \frac{1}{18,76635215} 2,718282^{-2,879619181} \\
 &= \frac{1}{18,76635215} \times 0,056156144 \\
 &= 0,002992385
 \end{aligned}$$

$$\begin{aligned}
 P(x_1 = |Tidak) &= \frac{1}{\sqrt{2.3,14.51,11164251}} 2,718282^{\frac{-(448-296)^2}{2(51,11164251)^2}} \\
 &= \frac{1}{\sqrt{320,981115}} 2,718282^{\frac{-(152)^2}{2(2612,4)}} \\
 &= \frac{1}{17,91594583} 2,718282^{\frac{-23104}{5224,8}} \\
 &= \frac{1}{17,91594583} 2,718282^{-4,421987444} \\
 &= \frac{1}{17,91594583} \times 0,012010339 \\
 &= 0,000670371
 \end{aligned}$$

Dari hasil perhitungan di atas diperoleh jika hasil umur panen 448 hari maka menghasilkan nilai $Ya = 0,002992385$ dan Tidak = 0,000670371.

2. Daerah Tanam (Ha)

$$\begin{aligned}
 P(x_2 = |Ya) &= \frac{1}{\sqrt{2.3,14.2,549298871}} 2,718282^{\frac{-(12-7,967741935)^2}{2(2,549298871)^2}} \\
 &= \frac{1}{\sqrt{16,00959691}} 2,718282^{\frac{-(4,032258065)^2}{2(6,498924731)}} \\
 &= \frac{1}{4,001199434} 2,718282^{\frac{-16,2591051}{12,99784946}} \\
 &= \frac{1}{4,001199434} 2,718282^{-1,250907325} \\
 &= \frac{1}{4,001199434} \times 0,286244962 \\
 &= 0,071539789
 \end{aligned}$$

$$\begin{aligned}
 P(x_2 = |Tidak) &= \frac{1}{\sqrt{2.3,14.2,593949003}} 2,718282^{\frac{-(12-7,857142857)^2}{2(2,593949003)^2}} \\
 &= \frac{1}{\sqrt{16,28999974}} 2,718282^{\frac{-(4,142857143)^2}{2(6,728571429)}} \\
 &= \frac{1}{4,036087182} 2,718282^{\frac{-17,16326531}{13,45714286}} \\
 &= \frac{1}{4,036087182} 2,718282^{-1,27540188} \\
 &= \frac{1}{4,036087182} \times 0,279318693 \\
 &= 0,069205317
 \end{aligned}$$

Dari hasil perhitungan di atas diperoleh jika hasil daerah tanam 12 hektar maka menghasilkan nilai Ya = 0,071539789 dan Tidak = 0,069205317.

3. Diameter Buah (Mm)

$$\begin{aligned}
 P(x_3 = |Ya) &= \frac{1}{\sqrt{2.3,14.14,25856392}} 2,718282^{\frac{-(95-91,57419355)^2}{2(14,25856392)^2}} \\
 &= \frac{1}{\sqrt{89,54378144}} 2,718282^{\frac{-(3,425806452)^2}{2(203,3066452)}} \\
 &= \frac{1}{9,462757602} 2,718282^{\frac{-11,73614984}{406,6132903}} \\
 &= \frac{1}{9,462757602} 2,718282^{-0,028863173} \\
 &= \frac{1}{9,462757602} \times 0,971549389 \\
 &= 0,102670852
 \end{aligned}$$

$$\begin{aligned}
 P(x_3 = |Tidak) &= \frac{1}{\sqrt{2.3,14.5,016933231}} 2,718282^{\frac{-(95-87,81904762)^2}{2(5,016933231)^2}} \\
 &= \frac{1}{\sqrt{31,50634069}} 2,718282^{\frac{-(7,180952381)^2}{2(25,16961905)}} \\
 &= \frac{1}{5,613050926} 2,718282^{\frac{-51,5660771}{50,3392381}} \\
 &= \frac{1}{5,613050926} 2,718282^{-1,024371426} \\
 &= \frac{1}{5,613050926} \times 0,359022067 \\
 &= 0,063962018
 \end{aligned}$$

Dari hasil perhitungan di atas diperoleh jika hasil diameter buah 95 mm maka menghasilkan nilai Ya = 0,102670852 dan Tidak = 0,063962018.

4. Bobot Buah (Kg)

$$\begin{aligned}
 P(x_4 = |Ya) &= \frac{1}{\sqrt{2.3,14.0,165443035}} 2,718282^{\frac{-(1,8-1,252258065)^2}{2(0,165443035)^2}} \\
 &= \frac{1}{\sqrt{1,03898226}} 2,718282^{\frac{-(0,547741935)^2}{2(0,027371398)^2}} \\
 &= \frac{1}{1,019304793} 2,718282^{\frac{-0,300021228}{0,054742796}} \\
 &= \frac{1}{1,019304793} 2,718282^{-5,480560941} \\
 &= \frac{1}{1,019304793} \times 0,004166992 \\
 &= 0,004088072
 \end{aligned}$$

$$\begin{aligned}
 P(x_4 = |Tidak) &= \frac{1}{\sqrt{2.3,14.0,195778057}} 2,718282^{\frac{-(1,8-0,902380952)^2}{2(0,195778057)^2}} \\
 &= \frac{1}{\sqrt{1,229486198}} 2,718282^{\frac{-(0,897619048)^2}{2(0,038329048)^2}} \\
 &= \frac{1}{1,108821987} 2,718282^{\frac{-0,805719955}{0,076658095}} \\
 &= \frac{1}{1,108821987} 2,718282^{-10,51056581} \\
 &= \frac{1}{1,108821987} \times 0,000027247 \\
 &= 0,000024573
 \end{aligned}$$

Dari hasil perhitungan di atas diperoleh jika hasil bobot buah 1,8 kg maka menghasilkan nilai Ya = 0,004088072 dan Tidak = 0,000024573.

5. Panjang Buah (Cm)

$$\begin{aligned}
 P(x_5 = |Ya) &= \frac{1}{\sqrt{2.3,14.1,863578475}} 2,718282^{\frac{-(25-17,13225806)^2}{2(1,863578475)^2}} \\
 &= \frac{1}{\sqrt{11,70327282}} 2,718282^{\frac{-(7,867741935)^2}{2(3,472924731)^2}} \\
 &= \frac{1}{3,421004651} 2,718282^{\frac{-61,90136316}{6,945849462}} \\
 &= \frac{1}{3,421004651} 2,718282^{-8,911993198} \\
 &= \frac{1}{3,421004651} \times 0,000134763 \\
 &= 0,000039393
 \end{aligned}$$

$$\begin{aligned}
P(x_5 = |\text{Tidak}) &= \frac{1}{\sqrt{2.3,14.0,195778057}} 2,718282^{\frac{-(25-0,902380952)^2}{2(0,195778057)^2}} \\
&= \frac{1}{\sqrt{13,00264068}} 2,718282^{\frac{-(10,47619048)^2}{2(4,286904762)}} \\
&= \frac{1}{3,505346986} 2,718282^{\frac{-109,7505669}{8,573809524}} \\
&= \frac{1}{3,505346986} 2,718282^{-14,46803145} \\
&= \frac{1}{3,505346986} 2,718282^{-14,46803145} \\
&= \frac{1}{3,505346986} \times 0,0000005207 \\
&= 0,0000001486
\end{aligned}$$

Dari hasil perhitungan di atas diperoleh jika hasil panjang 25 cm maka menghasilkan nilai Ya = 0,000039393 dan Tidak = 0,000000765.

6. Tinggi Tanaman (Cm)

$$\begin{aligned}
P(x_6 = |Ya) &= \frac{1}{\sqrt{2.3,14.17,25657113}} 2,718282^{\frac{-(97-89,5483871)^2}{2(17,25657113)^2}} \\
&= \frac{1}{\sqrt{108,3712667}} 2,718282^{\frac{-(7,451612903)^2}{2(297,7892473)}} \\
&= \frac{1}{10,4101521} 2,718282^{\frac{-55,52653486}{595,5784946}} \\
&= \frac{1}{10,4101521} 2,718282^{-0,093231262} \\
&= \frac{1}{10,4101521} \times 0,9109828 \\
&= 0,087509077
\end{aligned}$$

$$\begin{aligned}
P(x_6 = |\text{Tidak}) &= \frac{1}{\sqrt{2.3,14.19,14804478}} 2,718282^{\frac{-(97-88,61904762)^2}{2(19,14804478)^2}} \\
&= \frac{1}{\sqrt{120,2497212}} 2,718282^{\frac{-(8,380952381)^2}{2(366,647619)}} \\
&= \frac{1}{10,96584339} 2,718282^{\frac{-70,24036281}{733,2952381}} \\
&= \frac{1}{10,96584339} 2,718282^{-0,095787289} \\
&= \frac{1}{10,96584339} \times 0,908657277 \\
&= 0,082862507
\end{aligned}$$

Dari hasil perhitungan di atas diperoleh jika hasil tinggi tanaman 97 cm maka menghasilkan nilai Ya = 0,087509077 dan Tidak = 0,082862507.

7. Produksi Nanas (Kg)

$$\begin{aligned}
 P(x_7 = |Ya) &= \frac{1}{\sqrt{2.3,14.1850,328887}} 2,718282^{\frac{-(9000-5066,451613)^2}{2(1850,328887)^2}} \\
 &= \frac{1}{\sqrt{11620,06541}} 2,718282^{\frac{-(3933,548387)^2}{2(3423716,989)^2}} \\
 &= \frac{1}{107,7964072} 2,718282^{\frac{-15472802,91}{6847433,978}} \\
 &= \frac{1}{107,7964072} 2,718282^{-2,259649814} \\
 &= \frac{1}{107,7964072} \times 0,104387033 \\
 &= 0,000968372
 \end{aligned}$$

$$\begin{aligned}
 P(x_7 = |Tidak) &= \frac{1}{\sqrt{2.3,14.1820,603249}} 2,718282^{\frac{-(9000-4778,095238)^2}{2(1820,603249)^2}} \\
 &= \frac{1}{\sqrt{11433,3884}} 2,718282^{\frac{-(4221,904762)^2}{2(3314596,19)^2}} \\
 &= \frac{1}{106,9270237} 2,718282^{\frac{-17824479,82}{6629192,381}} \\
 &= \frac{1}{106,9270237} 2,718282^{-2,688786023} \\
 &= \frac{1}{106,9270237} \times 0,067963395 \\
 &= 0,000635605
 \end{aligned}$$

Dari hasil perhitungan di atas diperoleh jika hasil produksi nanas 9000 buah maka menghasilkan nilai $Ya = 0,000968372$ dan Tidak = $0,000635605$.

Tahap selanjutnya, yaitu: perhitungan metode *naive bayes* menggunakan rumus *likelihood*:

$$\begin{aligned}
 P(X|Ya) &= P(0,002992385|Ya) \times P(0,071539789|Ya) \times \\
 &\quad P(0,102670852|Ya) \times P(0,004088072|Ya) \times \\
 &\quad P(0,000039393|Ya) \times P(0,087509077|Ya) \times \\
 &\quad P(0,000968372|Ya) \times P(0,596|Ya) \\
 &= 1,78814 \times 10^{-16}
 \end{aligned}$$

$$\begin{aligned}
P(X|\text{Tidak}) &= P(0,000670371|\text{Tidak}) \times P(0,069205317|\text{Tidak}) \times \\
&\quad P(0,063962018|\text{Tidak}) \times P(0,000024573|\text{Tidak}) \times \\
&\quad P(0,000000765|\text{Tidak}) \times P(0,082862507|\text{Tidak}) \times \\
&\quad P(0,000635605|\text{Tidak}) \times P(0,404|\text{Tidak}) \\
&= 1,18663 \times 10^{-21}
\end{aligned}$$

4.4.3 Menghitung Probabilitas *Posterior*

Berdasarkan perhitungan *likelihood* yang telah diperoleh maka dapat diperoleh nilai probabilitas akhir adalah:

$$\begin{aligned}
P(X|Ya) &= \frac{1,78814 \times 10^{-16}}{(1,78814 \times 10^{-16} + 1,18663 \times 10^{-21})} \\
&= 0,999993364
\end{aligned}$$

$$\begin{aligned}
P(X|\text{Tidak}) &= \frac{1,18663 \times 10^{-21}}{(1,78814 \times 10^{-16} + 1,18663 \times 10^{-21})} \\
&= 0,000006636
\end{aligned}$$

4.4.4 Mencari Nilai *likelihood* Maksimum

Untuk mengkasifikasikan jenis nanas, apakah dapat dikatakan produktif atau tidak produktif. Melihat nilai akhir probabilitas yang hampir mendekati nilai 1 atau bernilai sama dengan 1. Dari hasil perhitungan probabilitas *posterior* diketahui bahwa hasil akhir yang diperoleh untuk setiap probabilitas tersebut, yaitu: untuk nilai akhir probabilitas Ya = 0,999993364 dan untuk nilai akhir probabilitas Tidak = 0,000006636. Sehingga dapat dikategorikan Ya yang memiliki arti nanas dengan atribut umur panen 448 hari, daerah tanam 12 ha, diameter buah 95 mm, bobot buah 1,8 kg, panjang buah 25 cm, tinggi tanman 97 cm, produksi nanas 9000 buah merupakan nanas PRODUKTIF.

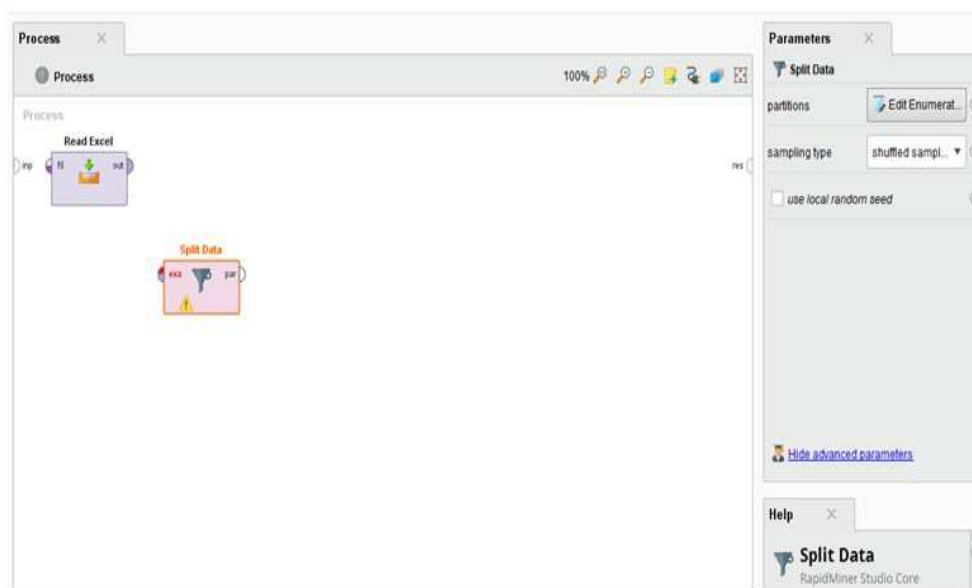
4.5 Analisa Algoritma *Naive Bayes* dengan *RapidMiner 7.1*

Tujuan utama pada penelitian ini untuk mendapatkan hasil klasifikasi nanas dengan menggunakan algoritma *naive bayes classification* pada Balai Penyuluhan Pertanian (BPP) Panai Tengah. Pengujian dilakukan untuk memprediksi nanas menggunakan *rapidminer 7.1*, nantinya akan diperoleh seberapa besar nilai akurasi metode *naive bayes* untuk prediksi nanas dengan menggunakan *data testing*.

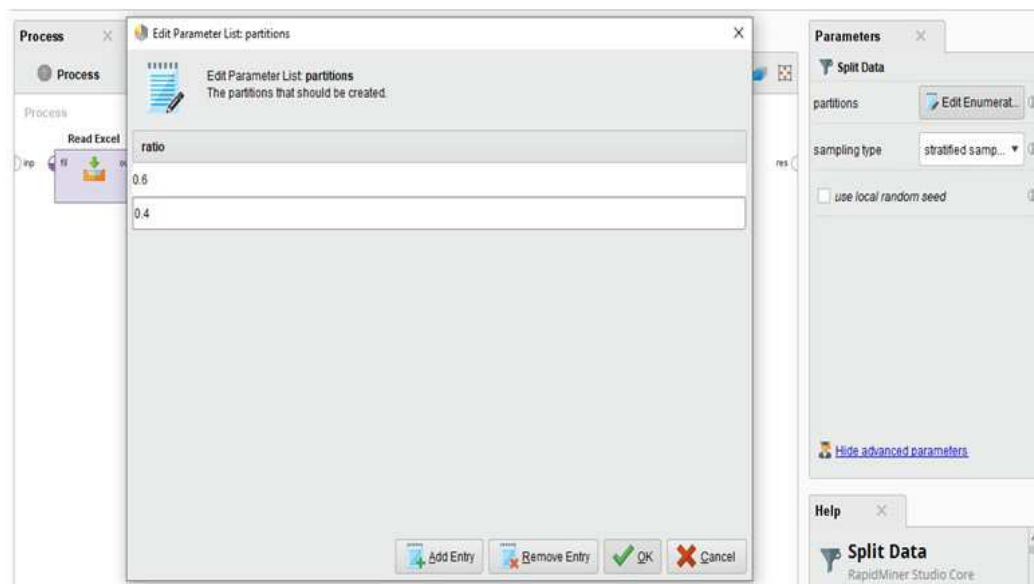
Sebelum menggunakan aplikasi *rapidminer 7.1*, hal pertama yang harus dilakukan adalah memeriksa terlebih dahulu data yg akan diolah. Data yang akan digunakan adalah *data set* nanas menggunakan metode *naive bayes*. Ketika seluruh data valid maka langkah selanjutnya pengolahan data memakai *rapidminer 7.1*. Berikut merupakan proses pengujian *naive bayes* menggunakan *rapidminer 7.1*.

4.5.1 Proses Split Data

Split Data digunakan untuk membagi *data set* menjadi dua bagian yaitu *data training* dan *data testing*, pada tampilan gambar 4.1 dibagian sebelah kanan atas ubah *sampling type* menjadi *stratified sampling* dan bagian *cpartitions* klik edit *enumeration* seperti gambar 4.9.



Gambar 4.1 Operator Read Excel dan Split Data



Gambar 4.2 Alur Split Data

Setelah mengubah *sampling type* menjadi *stratified sampling*, selanjutnya bagian *partitions* klik edit *enumeration* kemudian klik *add entry* untuk membagi *data set*. Ketik 0.6 sebagai *data training* dan 0.4 sebagai *data testing* setelah selesai maka klik OK.

Row No.	Kategori Pro...	Jenis Nanas	Umr Panen...	Daerah Tan...	Diameter Bu...	Bobot Buah ...	Panjang Bsa...	Tinggi Tana...	Produksi Na...	K	L
1	Ya	An1	252	8	90.300	1.350	17	99	6480	?	?
2	Ya	An3	252	8	99.100	1.400	20	91	5400	?	?
3	Ya	An4	280	10	96.100	1.550	22	74	8100	?	?
4	Tidak	An7	238	8	87.900	1.370	18	97	5400	?	?
5	Tidak	An12	238	8	84.900	0.980	15	109	8600	?	?
6	Ya	An13	238	8	86.600	1.250	17	96	6500	?	?
7	Tidak	An14	392	5	92.900	0.830	18	128	2700	?	?
8	Tidak	An16	336	8	79.200	0.740	10	73	4400	?	?
9	Tidak	An17	336	8	81.700	0.860	14	97	6500	?	?
10	Ya	An18	392	10	93.100	1.160	17.500	81	5400	?	?
11	Tidak	An19	280	10	87.800	0.900	13.500	96	5400	?	?
12	Ya	An20	280	12	93.300	1	16	100	6400	?	?
13	Ya	An22	252	12	91	1.500	21	125	6400	?	?
14	Tidak	An23	280	6	80.600	0.690	14	62	3000	?	?
15	Ya	An25	336	8	95.900	1.200	16	128	4300	?	?
16	Ya	An26	336	12	94.300	1.140	17	94	6400	?	?
17	Ya	An28	238	10	96.100	1.400	19	90	5400	?	?

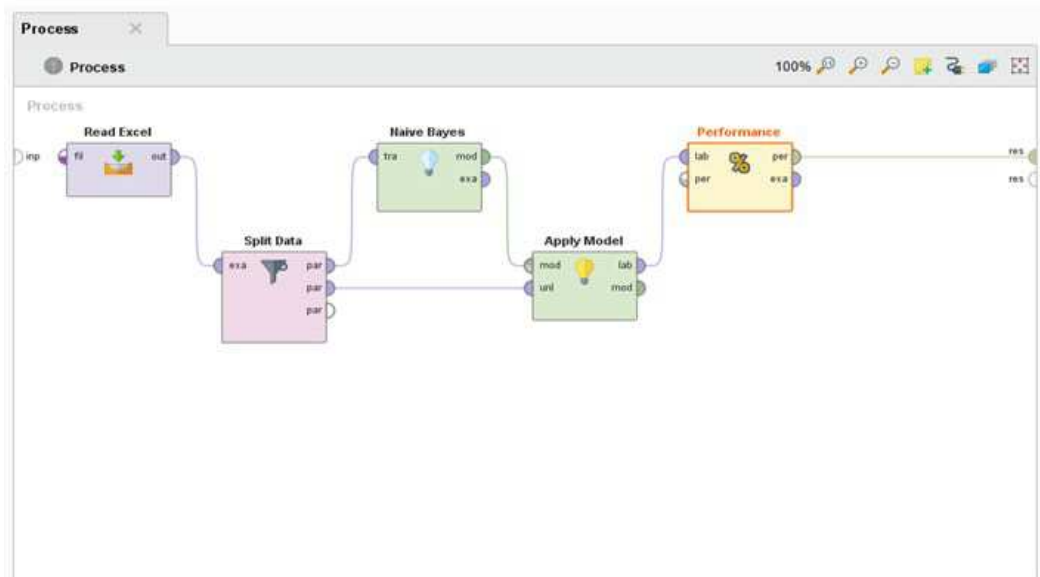
Gambar 4.3 ExampleSet Split Data

Hasil dari *split* data dapat dilihat pada gambar 4.3 diperoleh tabel exampleset dengan jumlah 52 data sebagai *data training*.

4.5.2 Akurasi Prediksi

Proses klasifikasi selanjutnya dengan *rapidminer* 7.1 menggunakan metode algoritma *naive bayes* pada data nanas. Membandingkan *data testing* dengan *data training* yang sudah diketahui *rule-rule* sebelumnya untuk mengetahui nilai akurasi prediksi. Menentukan berapa akurasi yang didapatkan menggunakan akurasi *naive bayes*, sebagai berikut :

1. Menambahkan operator *read excel* untuk memasukan *data set*.
2. Menambahkan operator *split* data yang digunakan sebagai pembagi *data set* menjadi *data training* dan *data testing*.
3. Menambahkan operator *naive bayes* yang digunakan sebagai metode dalam penelitian.
4. Menambahkan operator *apply* model untuk menguji *data testing* yang ingin ditentukan hasil prediksinya.
5. Menambahkan operator *performance* untuk mengetahui tingkat akurasi dengan metode *naive bayes*.



Gambar 4.4 Susunan Operator *Read Excel*, *Split Data*, *Naive Bayes*, *Apply Model*, dan *Performance*

Langkah terakhir adalah klik ikon *Run* pada *toolbar* untuk menampilkan hasilnya, seperti pada Gambar 4.6. Tunggu beberapa saat komputer untuk menyelesaikan perhitungannya.

Gambar 4.5 Icon *Run*

Setelah memperoleh hasil akurasi *performance vektor* dengan menggunakan *data training* dan *data testing* mendapat nilai akurasi sebesar 85,29%. Jumlah *data training* adalah 52 data, sebanyak 31 data yang dinyatakan nanas produktif “Ya”, dan sebanyak 21 data yang dinyatakan nanas tidak produktif “Tidak”. Pengujian berdasarkan pengukuran *confusion matrix* menghasilkan nilai *accuracy*, *precision*, dan *recall* yang cukup tinggi, sebagai berikut:

1. Accuracy

Dengan mengetahui jumlah data yang diklasifikasikan secara benar maka dapat diketahui akurasi hasil prediksinya 85,29% dari hasil *data testing*.

accuracy: 85.29%

	true Ya	true Tidak	class precision
pred. Ya	18	3	85.71%
pred. Tidak	2	11	84.62%
class recall	90.00%	78.57%	

Gambar 4.6 *Performance Accuracy*

Untuk menghitung akurasinya sebagai berikut:

$$\begin{aligned} Accuracy &= \frac{18+11}{18+11+3+2} \times 100\% \\ &= 85,29\% \end{aligned}$$

$$\begin{aligned} Error &= \frac{3+2}{18+11+3+2} \times 100\% \\ &= 14,70\% \end{aligned}$$

Dari perhitungan di atas dapat disimpulkan, bahwa klasifikasi dengan menggunakan metode *naive bayes* untuk menentukan klasifikasi nanas produktif atau tidak produktif. Menghasilkan tingkat akurasi sebesar 85,29% dan tingkat *error* 14,70%.

2. Precision

precision: 84.62% (positive class: Tidak)

	true Ya	true Tidak	class precision
pred. Ya	18	3	85.71%
pred. Tidak	2	11	84.62%
class recall	90.00%	78.57%	

Gambar 4.7 *Performance Precision*

$$\begin{aligned} \text{Precision} &= \frac{11}{11+2} \times 100\% \\ &= 84,62\% \end{aligned}$$

Dari perhitungan di atas dapat dilihat bahwa klasifikasi dengan menggunakan metode *naive bayes* menghasilkan *precision* sebesar 84,62%.

3. Recall

recall: 78.57% (positive class: Tidak)

	true Ya	true Tidak	class precision
pred. Ya	18	3	85.71%
pred. Tidak	2	11	84.62%
class recall	90.00%	78.57%	

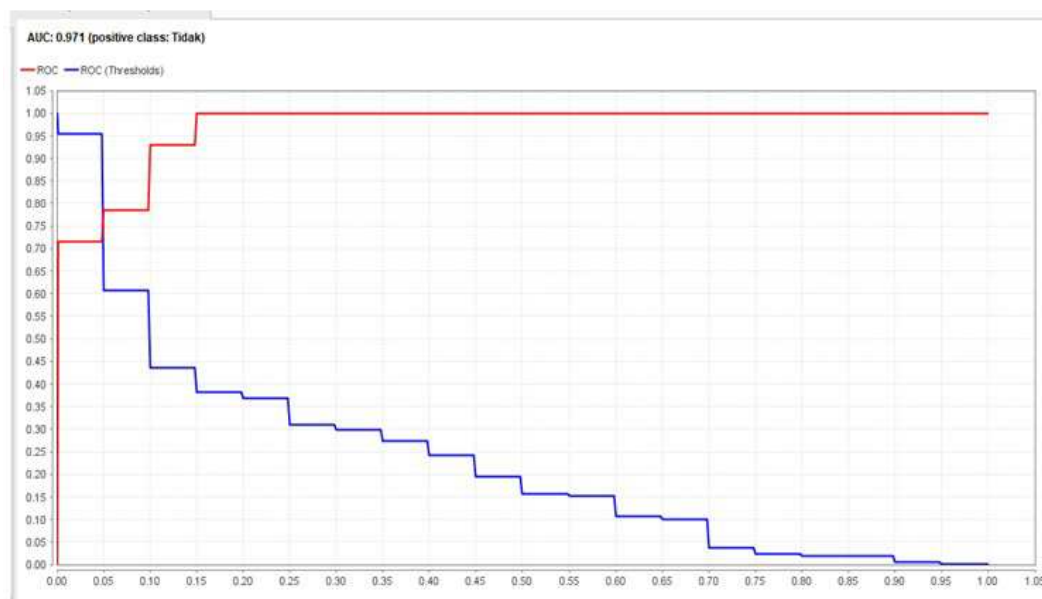
Gambar 4.8 *Performance Recall*

$$\begin{aligned} \text{Recall} &= \frac{11}{11+3} \times 100\% \\ &= 78,57\% \end{aligned}$$

Dari perhitungan di atas dapat dilihat bahwa klasifikasi dengan menggunakan metode *naive bayes* menghasilkan *precision* sebesar 78,57%.

4.5.3 Kurva ROC/AUC (Area Under Curve)

Pengujian berdasarkan pengukuran kurva ROC/ AUC sebesar 0.971 dengan tingkat akurasi *excellent classification*. Menggunakan metode algoritma *naive bayes*, proses pengambilan keputusan menjadi lebih mudah dan tepat sasaran karena memiliki tingkat akurasi yang cukup baik.



Gambar 4.9 Curva ROC

Kurva ROC digunakan untuk mengekspresikan data. Garis horizontal mewakili nilai *false positive* dan garis vertikal mewakili nilai *true positive*. Dari gambar di atas dapat diketahui bahwa nilai *area under curve* (AUC) model algoritma *naive bayes* 0.971, hal ini menunjukkan bahwa model algoritma *naive bayes* mencapai klasifikasi cukup sempurna.

4.6 Pembahasan

Penelitian yang telah dilakukan menghasilkan sebuah pola informasi dan pengetahuan baru, pada penggunaan proses *data mining* untuk memprediksi nanas di Panai Tengah. Penelitian ini menghasilkan suatu pola informasi dan pengetahuan yang sinkron. Menggunakan tujuan *data mining*, yaitu: pola *data training* dan *data testing* untuk mencari probabilitas setiap atribut yang memakai *data training* dan *data testing* untuk mendapatkan informasi baru, apakah pada data nanas tersebut termasuk dengan kategori produktif atau tidak produktif.

Adapun tujuan dari proses perhitungan *data mining* menggunakan *naive bayes* yaitu untuk mendapatkan hasil dari klasifikasi nanas pada Balai Penyuluhan Pertanian (BPP) Panai Tengah dengan memakai perhitungan manual dan pengujian menggunakan *software rapidminer 7.1* sebagai alat bantu untuk mempersentasikan taraf akurasi dari klasifikasi nanas tersebut.

Nilai *accuracy* yang diperoleh pada perhitungan manual dan pengujian menggunakan *rapidminer 7.1* adalah 85,29%. Untuk kelas “Ya” pada *rapidminer 7.1 prediction* sebesar 85,71% dan untuk kelas “Tidak” pada *rapidminer 7.1 prediction* sebesar 84,62% membuktikan bahwa metode *naive bayes* cukup baik digunakan untuk melakukan penjabaran prediksi nanas di Panai Tengah. Nilai 85,29% bisa disebabkan oleh beberapa kekurangan atribut data atau kompleksan data yang menyebabkan model dapat memprediksi nanas dengan akurat.

Sebanyak 34 data nanas dijadikan sebagai *data testing*, dengan menggunakan metode *naive bayes* ini ternyata dihasilkan penelitian bahwa nanas dengan kategori “Ya” sebanyak 20 nanas dan nanas dengan kategori “Tidak” sebanyak 14 nanas. Adapun karakter agronomi nanas yang mempengaruhi nanas tersebut produktif atau tidak produktif, yaitu: umur panen (hari), daerah tanam (ha), diameter buah (mm), bobot buah (kg), panjang buah (cm), tinggi tanaman (cm), dan produksi nanas dalam satu periode tanam (kg).

BAB 5

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil uji coba yang telah dilakukan pada penelitian ini, terbukti bahwa metode yang diusulkan dapat digunakan untuk mengklasifikasi nanas. Data keluaran yang dapat dihasilkan dan ditampilkan oleh aplikasi adalah suatu keputusan apakah nanas tersebut termasuk dalam klasifikasi jenis nanas yang produktif atau tidak produktif, sehingga dapat memberikan tolak ukur untuk Balai Penyuluhan Pertanian (BPP) Panai Tengah, industri nanas dan petani dalam membudidayakan nanas.

Pada penelitian ini *data mining* dan *naive bayes* mampu menampilkan berita prediksi nanas dengan membandingkan *data training* dengan *data testing* menggunakan aplikasi pendukung *rapidminer 7.1* diperoleh tingkat akurasi sebesar 85,29%, karena memiliki nilai lebih dari 50% ini berarti kinerja *naive bayes* yang diterapkan dalam permasalahan penelitian ini sudah cukup baik.

5.2 Saran

Untuk penelitian selanjutnya, peneliti menyarankan agar penelitian ini dikembangkan lagi dengan melakukan penambahan jumlah kriteria, agar pengujian pada akurasi selanjutnya dapat dinyatakan lebih akurat dari sebelumnya. Kemudian dalam penentuan jumlah *data training* dapat mempengaruhi hasil pengujian, karena pola *data training* yang dimaksud akan dijadikan sebagai *rule* dalam menentukan kelas pada *data testing*, Sehingga besar atau kecilnya persentase tingkat akurasi dipengaruhi juga oleh penentuan *data training*, maka dari itu diharapkan untuk penelitian selanjutnya dapat menambahkan jumlah *data training* lebih banyak lagi.

DAFTAR PUSTAKA

- Arifin, M.F, & Fitriana, D. (2018). Penerapan Algoritma Klasifikasi C4.5 Dalam Rekomendasi Penerimaan Mitra Penjualan Studi Kasus: PT Atria Artha Persada. *InComTech*. 8(2): 87102.
- Atika & Priatna. (2020). *Modul Perkuliahan Data Mining*. Jakarta: Universitas Bhayangkara Jakarta Raya.
- Budiyantara, A., Wijaya, A., & Gunawan, A. (2022). Analisis Data Mining Hotel Booking Menggunakan Model ID3. *Journal of Business and Audit Information Systems*, 4(1), 112.
- Das, Sanjiv. (2016). Data Science: Theories, Models, Algorithms, and Analytics. *In Sanjiv Ranjan Das*.
- Defiyanti, S. (2017). Integrasi Metode Clustering dan Klasifikasi untuk Data Numerik. *Citee*, 256261.
- Hadiyantono, Tane. (2018). Kemtan dorong produktivitas dan ekspor nanas. <https://industri.kontan.co.id/news/kemtan-dorong-produktivitas-dan-ekspor-nanas>. (16 Oktober 2022).
- Ishak, M., Parwati, W., & Wirianata, H. (2017). Kajian Tumpang Sari Nanas Pada Berbagai Umur TBM Kelapa Sawit Di Perkebunan Rakyat. *Jurnal Agromast*, 2(1).
- Kadafi, A. (2018). Perbandingan Algoritma Klasifikasi Untuk Penjurusan Siswa SMA. *Jurnal Eltikom*, 2(2), 6777.
- Kurniawan, Y. (2018). Perbandingan Algoritma Naive Bayes dan C.45 dalam Klasifikasi Data Mining. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 5(4), 455464.
- Luvia, Y. S., Windarto, A. P., Solikhun, S., & Hartama, D. (2017). Penerapan Algoritma C4.5 Untuk Klasifikasi Predikat Keberhasilan Mahasiswa Di Amik Tunas Bangsa. *Jurasik (Jurnal Riset Sistem Informasi Dan Teknik Informatika)*, 1(1), 7579.
- Mayadewi, P., & Rosely, E. (2015). Prediksi Nilai Proyek Akhir Mahasiswa Menggunakan Algoritma Klasifikasi Data Mining. *Seminar Nasional Sistem Informasi Indonesia*, 329334.
- Muhamad, H., Prasojo, C., Sugianto, N., Surtiningsih, L., & Cholissodin, I. (2017). Optimasi Nave Bayes Classifier Dengan Menggunakan Particle Swarm Optimization Pada Data Iris. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 4(3), 180184.
- Partogi Yosua. (2021). Perancangan Metode Decision Tree Terhadap Sistem Perpustakaan Stmik Kuwera. *Jurnal Sistem Informasi Dan Teknologi (SINTEK)*, 1(2), 2026.
- Putri, A. B., & Anita. (2017). Efek Anti Inflamasi Enzim Bromelin Nanas terhadap Osteoarthritis. *Jurnal Kesehatan*, 8(3), 489500.
- Raschka, S. (2014). *Naive Bayes and Text Classification I - Introduction and Theory*(pp.1-20).

- Rizal Syaiful. (2021). Kemuliaan Menjadi Petani dalam Islam. <https://iaiq.ac.id/blog/kemuliaan-menjadi-petani-dalam-islam/>. (11 Maret 2022).
- Sabransyah, M., Nasution, Y., & Amijaya, F. (2017). Aplikasi Metode Naive Bayes dalam Prediksi Risiko Penyakit Jantung. *Jurnal Eksponensial*, 8(2), 111118.
- Saputro, I., & Sari, B. (2019). Uji Performa Algoritma Nave Bayes untuk Prediksi Masa Studi Mahasiswa. *Citec Journal*, 6(1), 111.
- Sartika, D., & Indra, D. (2017). Perbandingan Algoritma Klasifikasi Naive Bayes, Nearest Neighbour, dan Decision Tree pada Studi Kasus Pengambilan Keputusan Pemilihan Pola Pakaian. *Jurnal Teknik Informatika Dan Sistem Informasi*, 1(2), 151161.
- Selvia Aprilyanti. (2018). Pengaruh konsentrasi NaOH dan waktu hidrolisis terhadap kadar selulosa pada daun nanas. *Jurnal Teknik Kimia*, 24(1), 2831.
- Setiawan, H. R. (2017). Kontribusi al-khawarizmi dalam perkembangan ilmu astronomi. *Al-Marshad: Jurnal Astronomi Dan Ilmu Berkaitan*, Vol. 1(No. 1), 6891.
- Suryono, W., Saptono, Ristu, & Wiranto, W. (2017). Implementasi Pengembangan Smart Helpdesk di UPT TIK UNS Menggunakan Algoritma Naive Bayes Classifier. *Seminar Nasional Aplikasi Teknologi Informasi (SNATI)*, 3943.
- Tiara, F., Surtono, A., & Supriyanto, A. (2020). Rancang Bangun Purwarupa Pemilah Nanas Berdasarkan Tingkat Kematangan Menggunakan Mikrokontroler Blue Pill STM32F103C8T6. *Journal Of Energy, Material, and Instrumentation Technology*, 1(3), 8190.
- Wulandari, A., & Daningsih, E. (2019). *Modifikasi Biscotti dengan Cita Nanas*. 08(02), 193206.
- Yudhistira, A., Suroso, & Sholihin. (2022). *Sistem Pemantauan Penggunaan Protokol Kesehatan COVID-19 Menggunakan Metode Haar Cascade Dan Neural Network*. 11(2), 3246.

LAMPIRAN
DATA SET ATRIBUT NANAS PANAI TENGAH

No	Jenis Nanas	Umur Panen (hari)	Daerah Tanam (rante)	Diameter Buah (mm)	Bobot Buah (kg)	Panjang Buah (cm)	Tinggi Tanaman (cm)	Produksi Nanas (buah)	Kategori Produktif
1	An1	252	8	90,3	1,35	17	99	6480	Ya
2	An2	252	8	98,1	1,7	23	110	5500	Ya
3	An3	252	8	99,1	1,4	20	91	5400	Ya
4	An4	280	10	96,1	1,55	22	74	8100	Ya
5	An5	280	10	92,7	0,95	14	94	4300	Tidak
6	An6	238	6	84,4	0,8	13	79	5400	Tidak
7	An7	238	8	87,9	1,37	18	97	5400	Tidak
8	An8	252	10	87,5	0,89	13,5	89	5400	Tidak
9	An9	280	8	85,3	1	14,5	119	6480	Ya
10	An10	280	8	89,1	0,86	15	103	7600	Tidak
11	An11	280	8	89,1	1,2	18	103	6400	Ya
12	An12	238	8	84,9	0,98	15	109	8600	Tidak
13	An13	238	8	86,6	1,25	17	96	6500	Ya
14	An14	392	5	92,9	0,83	18	128	2700	Tidak
15	An15	336	12	93,8	1,25	18,5	101	8600	Ya
16	An16	336	8	79,2	0,74	10	73	4400	Tidak
17	An17	336	8	81,7	0,86	14	97	6500	Tidak
18	An18	392	10	93,1	1,16	17,5	81	5400	Ya
19	An19	280	10	87,8	0,9	13,5	96	5400	Tidak
20	An20	280	12	93,3	1	16	100	6400	Ya
21	An21	280	12	92,1	1,35	19	80	6400	Ya
22	An22	252	12	91	1,5	21	125	6400	Ya
23	An23	280	6	80,6	0,69	14	62	3000	Tidak
24	An24	336	10	91,2	1,2	15	93	5400	Ya
25	An25	336	8	95,8	1,2	16	128	4300	Ya
26	An26	336	12	94,3	1,14	17	94	6400	Ya
27	An27	238	10	93,7	0,12	14	141	8100	Tidak
28	An28	238	10	96,1	1,4	19	90	5400	Ya
29	An29	238	12	90,7	1,14	16,5	62	6480	Ya
30	An30	392	12	94,9	1,2	21,5	117	6480	Ya
31	An31	280	8	82,2	0,68	12	99	4300	Tidak
32	An32	392	10	93	1	16	99	5400	Ya
33	An33	392	10	89,9	0,98	13	67	5400	Tidak

No	Jenis Nanas	Umur Panen (hari)	Daerah Tanam (rante)	Diameter Buah (mm)	Bobot Buah (kg)	Panjang Buah (cm)	Tinggi Tanaman (cm)	Produksi Nanas (buah)	Kategori Produktif
34	An34	252	12	92,3	1,1	15,5	90	7800	Tidak
35	An35	252	6	94,2	0,86	14	110	3100	Tidak
36	An36	252	10	92,5	0,95	14	101	5400	Tidak
37	An37	336	10	92,8	1,37	17,5	91	5400	Tidak
38	An38	336	8	92,6	1,2	15	90	4400	Ya
39	An39	336	8	95,2	1,15	15	67	4000	Ya
40	An40	336	8	98,9	1,15	15,5	82	4320	Ya
41	An41	238	4	90	0,9	16	99	3000	Tidak
42	An42	392	4	93,8	1,14	16,5	71	3000	Ya
43	An43	392	6	89,9	1,1	16	85	3200	Tidak
44	An44	392	6	91,7	1,35	18	93	3200	Ya
45	An45	392	6	16,5	1,25	16,5	67	3200	Ya
46	An46	280	5	84,3	0,8	13	112	2700	Tidak
47	An47	252	12	97,7	1,2	14	109	7400	Ya
48	An48	252	6	90,9	1,25	18	67	3240	Ya
49	An49	252	10	95,5	1,23	16,5	99	5400	Ya
50	An50	280	10	89,2	1,14	16,5	99	5400	Ya
51	An51	252	12	92,6	1	15,5	82	6400	Tidak
52	An52	280	12	87,9	0,9	18	67	6400	Tidak
53	An53	280	8	87,6	0,94	13,5	62	7500	Tidak
54	An54	280	12	92,4	1,2	15,5	87	6400	Ya
55	An55	280	4	85,2	0,7	14	62	3000	Tidak
56	An56	336	5	91,8	1,24	18	99	2700	Ya
57	An57	252	5	89,9	0,95	15,5	101	2700	Tidak
58	An58	336	4	82	0,75	13,5	110	2160	Tidak
59	An59	252	10	93,3	1,16	16	99	5400	Ya
60	An60	252	10	94,7	1,15	16	67	5400	Tidak
61	An61	252	12	87,4	0,7	13,5	87	6480	Tidak
62	An62	392	6	93,5	1,18	15,5	99	3240	Ya
63	An63	392	6	99,6	1,6	19,5	101	3240	Ya
64	An64	392	8	92,3	1,5	14	128	4320	Ya
65	An65	392	5	85	0,73	13	93	2700	Tidak
66	An66	252	5	92	1,35	21	99	2700	Ya
67	An67	252	12	95	1,15	16	103	6400	Ya
68	An68	336	4	99,7	1,25	15,5	67	2500	Ya
69	An69	252	5	94,2	1,28	19	99	2700	Ya

No	Jenis Nanas	Umur Panen (hari)	Daerah Tanam (rante)	Diameter Buah (mm)	Bobot Buah (kg)	Panjang Buah (cm)	Tinggi Tanaman (cm)	Produksi Nanas (buah)	Kategori Produktif
70	An70	252	6	95,1	1,1	17,5	87	3200	Ya
71	An71	252	8	92,2	1,26	17	67	7600	Ya
72	An72	252	6	78,4	0,71	13	128	3300	Tidak
73	An73	392	12	95,5	1,1	15	65	5400	Ya
74	An74	252	6	92,9	0,92	16	67	3400	Tidak
75	An75	392	8	88	1,23	16	99	4300	Ya
76	An76	336	4	93,8	1,8	15,8	112	2100	Ya
77	An77	336	4	98,1	1,1	14,5	101	2100	Tidak
78	An78	336	6	96,1	1,1	15	110	8400	Ya
79	An79	336	8	92,4	1,15	16,5	67	8500	Ya
80	An80	252	4	89,3	0,7	13	110	2000	Tidak
81	An81	252	8	94,2	1,28	17	67	4400	Ya
82	An82	280	8	99,4	1,19	16,8	87	4400	Ya
83	An83	392	5	91,8	1,39	18,5	67	2700	Ya
84	An84	392	6	97,8	1,39	19	86	6600	Ya
85	An85	280	5	97	1,1	15	101	2700	Tidak
86	An86	252	4	92,6	1,23	17	99	2160	Ya

HASIL DATA TESTING DENGAN RAPIDMINER

No	Jenis Nanas	Umur Panen (hari)	Daerah Tanam (rante)	Diameter Buah (mm)	Bobot Buah (kg)	Panjang Buah (cm)	Tinggi Tanaman (cm)	Produksi Nanas (buah)	Kategori Produktif
1	An2	252	8	98,1	1,7	23	110	5500	Ya
2	An5	280	10	92,7	0,95	14	94	4300	Tidak
3	An6	238	6	84,4	0,8	13	79	5400	Tidak
4	An8	252	10	87,5	0,89	13,5	89	5400	Tidak
5	An9	280	8	85,3	1	14,5	119	6480	Ya
6	An10	280	8	89,1	0,86	15	103	7600	Tidak
7	An11	280	8	89,1	1,2	18	103	6400	Ya
8	An15	336	12	93,8	1,25	18,5	101	8600	Ya
9	An21	280	12	92,1	1,35	19	80	6400	Ya
10	An24	336	10	91,2	1,2	15	93	5400	Ya
11	An27	238	10	93,7	0,12	14	141	8100	Tidak
12	An29	238	12	90,7	1,14	16,5	62	6480	Ya
13	An30	392	12	94,9	1,2	21,5	117	6480	Ya
14	An34	252	12	92,3	1,1	15,5	90	7800	Tidak
15	An41	238	4	90	0,9	16	99	3000	Tidak
16	An43	392	6	89,9	1,1	16	85	3200	Tidak
17	An44	392	6	91,7	1,35	18	93	3200	Ya
18	An46	280	5	84,3	0,8	13	112	2700	Tidak
19	An47	252	12	97,7	1,2	14	109	7400	Ya
20	An48	252	6	90,9	1,25	18	67	3240	Ya
21	An54	280	12	92,4	1,2	15,5	87	6400	Ya
22	An57	252	5	89,9	0,95	15,5	101	2700	Tidak
23	An60	252	10	94,7	1,15	16	67	5400	Tidak
24	An62	392	6	93,5	1,18	15,5	99	3240	Ya
25	An63	392	6	99,6	1,6	19,5	101	3240	Ya
26	An64	392	8	92,3	1,5	14	128	4320	Ya
27	An67	252	12	95	1,15	16	103	6400	Ya
28	An69	252	5	94,2	1,28	19	99	2700	Ya
29	An70	252	6	95,1	1,1	17,5	87	3200	Ya
30	An72	252	6	78,4	0,71	13	128	3300	Tidak
31	An75	392	8	88	1,23	16	99	4300	Ya
32	An77	336	4	98,1	1,1	14,5	101	2100	Tidak
33	An80	252	4	89,3	0,7	13	110	2000	Tidak
34	An81	252	8	94,2	1,28	17	67	4400	Ya