

Terbit online pada laman : <http://teknosi.fti.unand.ac.id/>

## Jurnal Nasional Teknologi dan Sistem Informasi

| ISSN (Print) 2460-3465 | ISSN (Online) 2476-8812 |



# Implementasi Sistem Klasifikasi Tim Kerja Menggunakan Algoritma *Naïve Bayes* (Studi Kasus: PT.MNL)

Tasya Angelya<sup>a</sup>, Abdul Rahman<sup>b</sup>, Iis Pradesan<sup>c</sup>

<sup>a,b,c</sup> Sistem Informasi, Universitas Multi Data Palembang, Palembang, 30113, Indonesia

### INFORMASI ARTIKEL

*Sejarah Artikel:*

Diterima Redaksi: 15 Maret 2023

Revisi Akhir: 24 Agustus 2023

Diterbitkan Online: 31 Agustus 2023

### KATA KUNCI

*CRISP-DM,  
Naive Bayes,  
Klasifikasi Tim,  
Hasil Kerja*

### KORESPONDENSI

E-mail: [tasyaangelya12@mhs.mdp.ac.id](mailto:tasyaangelya12@mhs.mdp.ac.id)

### ABSTRACT

PT.MNL adalah perusahaan swasta yang bergerak dibidang jasa kehutanan khususnya Hutan Tanaman Industri. Perusahaan memiliki banyak areal kerja yang memiliki banyak tim kerja di semua arealnya. Mengingat besarnya tim kerja, pemilik perusahaan harus bisa menjaga kinerja tim kerjanya agar selalu berada dalam posisi baik. Dengan semakin banyaknya tim kerja maka data yang masuk ke dalam perusahaan semakin banyak pula, data-data tersebut tentunya perlu dimanfaatkan dengan cara mengolah dan menganalisis untuk menghasilkan suatu informasi yang berguna bagi perusahaan. Sehingga dapat digunakan dalam mengklasifikasikan seluruh anggota-anggota tim pada lokasi kerja tertentu. Maka dari itu, diperlukan sebuah teknologi *Data Mining* yang dapat membantu perusahaan dalam memilih nama tim yang sebaiknya dipilih. Metode yang dipakai untuk klasifikasi tim kerja adalah algoritma *Naïve Bayes*. Penelitian ini menghasilkan tingkat akurasi klasifikasi sebesar 93,66%, hasil recall 90,74% serta hasil precision sebesar 88,58% sehingga hal ini menunjukkan bahwa sistem klasifikasi tim menggunakan algoritma *Naive Bayes* dapat digunakan oleh pemilik perusahaan sebagai bahan pertimbangan dalam mengatur persebaran tim di setiap lokasi kerja.

## 1. PENDAHULUAN

Pertumbuhan teknologi semakin cepat dan maju dari waktu ke waktu sehingga dibutuhkan keahlian-keahlian baru bagi manusia dalam menjalankan bisnis, salah satunya keahlian dalam mengolah data yang berperan penting dalam mendukung rencana bisnis. Perusahaan tidak dapat lagi bersaing secara kompetitif jika hanya memanfaatkan sistem operasional yang ada karena sistem operasional saja tidak cukup untuk menghasilkan informasi strategis dalam mendukung pertumbuhan bisnis [1].

Dalam menjalankan kegiatan operasional perusahaan akan menghasilkan data yang banyak sehingga akan menimbulkan bertumpuknya data [2]. Maka dari itu, dibutuhkanlah sebuah teknologi yang dapat digunakan dalam memperoleh serta menggali pengetahuan-pengetahuan baru dalam data yang tidak dapat diketahui secara kasat mata sehingga dapat membantu perusahaan dalam mengatur rencana bisnisnya karena data

merupakan harta perusahaan yang sangat penting dan berharga [2].

PT.MNL merupakan salah satu perusahaan jasa yang berdiri sejak tahun 2016. Perusahaan ini bergerak dibidang kehutanan khususnya hutan tanaman industri. PT. MNL merupakan sub kontraktor dari group perusahaan besar yang bergerak dibidang pulp dan kertas, agribisnis dan pangan,. PT. MNL memiliki banyak tim kerja yang tersebar disemua lokasi Sinarmas Group khususnya di Sumatera Selatan. Semua hal yang berkaitan dengan tenaga kerja menjadi tanggung jawab dari PT.MNL. PT. MNL memiliki tim kerja yang terdiri dari 130-an tim. Setiap tim terdiri dari sub-sub tim yang masing-masing memiliki 5-7 orang anggota. Di dalam lokasi, tim mengerjakan proyek-proyek kehutanan seperti langsir pupuk, langsir bibit, muat pupuk, bongkar pupuk dan sebagainya.

Mengingat besarnya tim kerja, pemilik perusahaan harus bisa menjaga kinerja tim kerjanya agar selalu baik. Selama ini pihak perusahaan dalam menempatkan tim kerja dilakukan dengan cara

manual tanpa melihat riwayat data sebelumnya sehingga cara ini belum maksimal dalam menempatkan tim kerjanya. Selain itu, saat ini PT.MNL belum memanfaatkan teknologi informasi untuk membantu kegiatan operasionalnya sehingga data transaksi yang setiap hari dilaporkan akan di input satu persatu ke dalam *Microsoft Excel* yang sudah mencapai 168.000 baris data. Data-data tersebut tidak dimanfaatkan perusahaan sehingga hanya menumpuk saja dari awal perusahaan berdiri sampai sekarang. Maka dari itu dibutuhkan sebuah cara dalam menganalisis serta mengelola data untuk mendapatkan informasi yang akurat selaras dengan pertumbuhan data yang cepat.

Untuk mengatasi masalah tersebut maka dibutuhkan teknologi *Data Mining* [3]. *Data Mining* adalah proses yang digunakan untuk mengolah data yang besar menjadi informasi yang berguna. Menurut [4], *Data Mining* adalah kumpulan proses yang dapat mengekstraksi nilai tambah dari kumpulan data berupa informasi yang sering digunakan secara manual untuk tujuan klasifikasi, estimasi, prediksi, asosiasi maupun pengklasteran.

Untuk mendapatkan informasi mengenai pembagian tim kerja yang efektif, diperlukan pengolahan data yang dapat dilakukan dengan menggunakan metode klasifikasi *Data Mining*. Metode klasifikasi adalah proses sistematis dalam membangun klasifikasi dari data yang diinput. Setiap metode menggunakan algoritma pembelajaran untuk mendapatkan suatu model yang paling sesuai dengan hubungan antara atribut dan label kelas pada data yang diinput [5][6].

Teknologi *Data Mining* telah banyak digunakan dalam sistem informasi untuk pengambilan keputusan manajemen. Metode klasifikasi *Data Mining* yang digunakan untuk memprediksi, diantaranya *Naïve Bayes Classifier*, *Decision Tree*, *Neural Network* dan *K-Nearest Neighbour*. Salah satu algoritma yang terkenal dalam klasifikasi adalah *Naive Bayes*. Menurut [7], keuntungan menggunakan algoritma *Naive Bayes* adalah metode ini hanya memerlukan sedikit data dalam menentukan estimasi parameter yang diperlukan untuk sistem klasifikasi. *Naive Bayes* bekerja dengan baik dalam situasi dunia nyata yang lebih kompleks dari pada yang diharapkan [8]. Menurut [9], klasifikasi *Naive Bayes* merupakan algoritma klasifikasi yang sederhana yang kinerjanya sebanding dengan klasifikasi *Decision Tree* dan *Neural Network*.

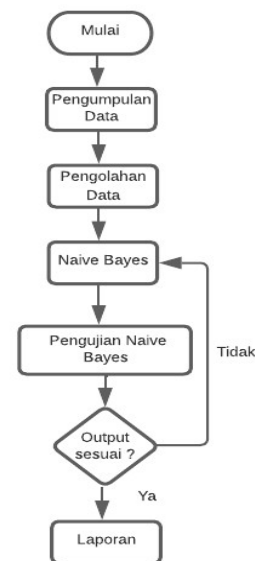
Keefektifan dan kesederhanaan dari algoritma *Naive Bayes* membuat algoritma ini telah banyak diaplikasikan pada berbagai bidang, seperti penelitian yang dilakukan oleh [10] mengenai klasifikasi untuk data pelanggan menyatakan bahwa klasifikasi menggunakan *Naive Bayes* dapat membantu pemilik perusahaan dalam memberikan bonus sehingga dapat meningkatkan kualitas pelayanan yang lebih baik lagi kepada pelanggannya. Penelitian terkait lainnya terkait prediksi kebakaran hutan menggunakan algoritma *Naive Bayes* menyatakan bahwa klasifikasi menggunakan algoritma *Naive Bayes* mempunyai tahapan perhitungan yang singkat dan memiliki nilai akurasi yang tinggi baik pada data berukuran kecil maupun data yang besar serta algoritma ini juga lebih mudah digunakan dibandingkan dengan algoritma lain[11]. Algoritma naïve bayes juga digunakan dalam melakukan klasifikasi dokumen berita[12]. Algoritma *Naive Bayes* banyak diimplementasikan untuk sistem pengambilan keputusan, diantaranya: sistem pengambilan keputusan untuk

menganalisis gangguan perkembangan anak dan memberikan rekomendasi dalam mengambil keputusan untuk mendeteksi gangguan perkembangan pada anak [13], pemilihan mata kuliah pilihan[14], pemberian bantuan sosial[15], sistem seleksi penerimaan dan penempatan sumber daya manusia pada suatu perusahaan[16], dan sistem pemberian diskon dalam sistem multi-kriteria[17].

Pada kajian penelitian-penelitian terkait yang didapat, label pada klasifikasi *Naïve Bayes* tidak begitu banyak berkisar 3-5 label. Pada penelitian ini diterapkan *Data Mining* pada dataset hasil kerja dengan pemodelan *supervised learning* menggunakan algoritma *Naive Bayes* dimana tim kerja dijadikan label yang jumlahnya 80 tim kerja. Manfaat dari penelitian ini adalah sebagai bahan pertimbangan perusahaan dalam pengambilan keputusan ketika mengklasifikasikan tim kerja sehingga perusahaan bisa mengatur persebaran tim pada setiap lokasi dengan harapan agar perusahaan memperoleh keuntungan dan manfaat seperti yang diharapkan.

## 2. METODE

Tahapan yang dilakukan dalam penelitian ini digambarkan dalam diagram alir pada Gambar 1.



Gambar 1. Tahapan Penelitian

Pada gambar 1, tahapan yang dilakukan dalam penelitian ini sebagai berikut:

### 2.1. Pengumpulan Data

Dalam memperoleh data perlu dilakukan penentuan sumber data yang tepat sehingga relevan dengan penelitian yang akan dilakukan. Sumber data yang digunakan adalah data hasil kerja tim pada PT.MNL dari tahun 2019 hingga 2022. Adapun metode pengumpulan data adalah sebagai berikut:

#### a. Wawancara

Proses ini dilakukan melalui wawancara langsung dengan pihak perusahaan untuk mendapatkan penjelasan tentang sistem

yang berjalan dan serta untuk mendapatkan data yang objektif tentang sistem penyimpanan data dan pengelolaan data di perusahaan.

b. Studi literatur

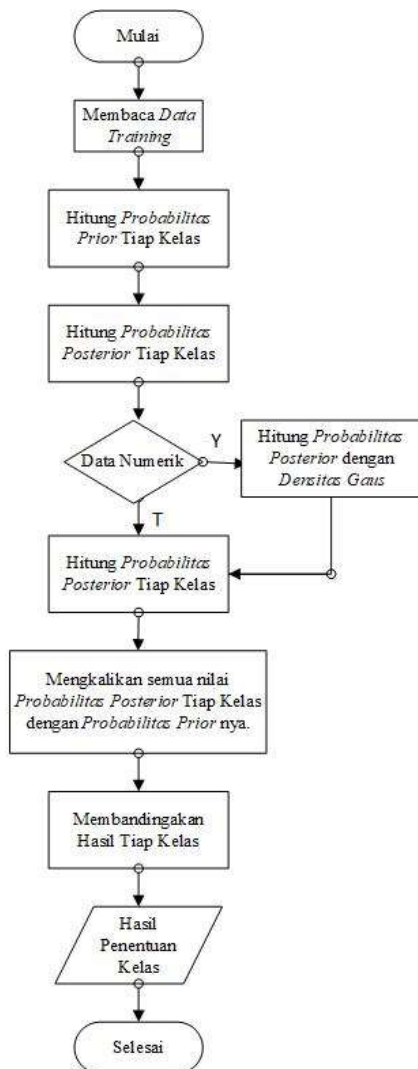
Proses ini dilakukan pengumpulan data dengan menggali informasi dari data yang diperbolehkan oleh perusahaan dalam menunjang penelitian dan jurnal ilmiah terkait.

2.2. **Pengolahan Data**

Pada tahap ini dilakukan pengolahan data untuk memperoleh data yang telah siap digunakan. Pengolahan data dimulai dari proses membersihkan data dari data yang tidak lengkap, data yang *noisy* dan data yang terdapat *redundancy*.

2.3. **Naive Bayes**

Tahapan implementasi *Data Mining* dengan menggunakan algoritma *Naive Bayes* yang dapat dilihat pada gambar 2.



Gambar 2. Alur Algoritma *Naive Bayes*

Persamaan (1) menunjukkan perhitungan pada teorema *Naive Bayes*:

$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(X)} \dots \dots (1)$$

Penjelasan dari persamaan (1) sebagai berikut:

$X$  : Data dengan class yang belum diketahui.

$H$  : Hipotesis pada data  $X$  yang merupakan suatu class khusus.

$P(H|X)$  : Nilai probabilitas pada hipotesis  $H$  berdasarkan kondisi  $X$  (Probabilitas Posterior).

$P(H)$  : Nilai probabilitas pada hipotesis  $H$  (Probabilitas Prior).

$P(X|H)$  : Nilai probabilitas  $X$  yang berdasarkan dengan kondisi  $H$ .

$P(X)$  : Nilai probabilitas pada  $X$ .

2.4. **Pengujian Naive Bayes**

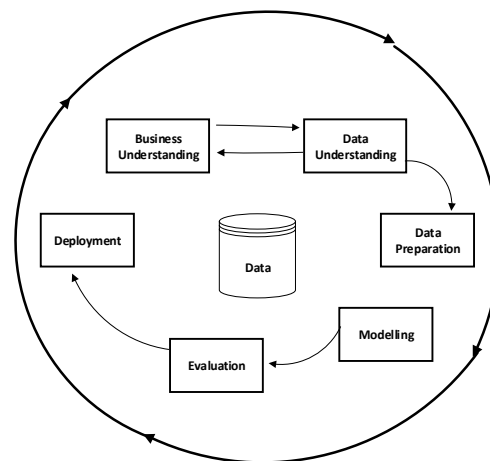
Pada tahap ini dilakukan pengujian algoritma *Naive Bayes* dengan melakukan *training* serta *testing* pada data dengan menggunakan *Naive Bayes*. Setelah itu membandingkan apakah *output* yang dihasilkan telah sesuai..

2.5. **Pengolahan Data**

Pada tahap ini dilakukan pengolahan data untuk memperoleh data yang telah siap digunakan. Pengolahan data dimulai dari proses membersihkan data dari data yang tidak lengkap, data yang *noisy* dan data yang terdapat *redundancy*.

2.6. **Metode Cross Industry Standard Process for Data Mining**

Metodologi *Cross Industry Standard Process for Data Mining* (CRISP-DM) terdiri dari tahapan: *business understanding*, *data understanding*, *modeling*, *evaluation*, dan *deployment* [18]. Metodologi CRISP-DM ditunjukkan pada Gambar 3.



Gambar 3. Metode CRISP-DM [19]

Metode CRISP-DM dijelaskan sebagai berikut:

1) *Business Understanding*

Pada tahap *business understanding*, berfokus pada pemahaman tujuan kebutuhan berdasarkan penilaian bisnis. Kemudian pemahaman tersebut diubah menjadi sebuah rencana awal *Data Mining* yang dirancang untuk mencapai tujuan. Pemahaman bisnis mengacu pada aturan penempatan tim di lokasi kerja yang ditentukan oleh PT. MNL. Pada tahap ini diperlukan pemahaman tentang latar belakang dan tujuan penempatan masing-masing tim di lokasi.

### 2) Data Understanding

Pada fase ini dilakukan pemahaman data berdasarkan data hasil kerja. Fase ini digunakan untuk memahami bentuk data, format data serta kegunaan data secara mendalam. Data-data tersebut dijelaskan sebagai berikut:

- a. Tabel Dimensi Pekerjaan menjelaskan mengenai informasi tentang jenis-jenis kegiatan yang dilakukan. Tabel ini terdiri dari No\_SPK\_Sistem, No\_SPK\_Manual, Id\_Pekerjaan, dan Nama\_Pekerjaan
- b. Tabel Dimensi Lokasi menjelaskan mengenai lokasi kerja yang dilakukan tim. Tabel ini berisi Id\_Petak, No\_Petak, Lokasi, Blok dan Nama\_PT
- c. Tabel Dimensi Waktu menjelaskan mengenai informasi tentang skala waktu, yaitu tanggal, bulan dan tahun. Tabel ini berisi Id\_Waktu dan Detail\_tgl
- d. Tabel Dimensi Tem berisi mengenai nama-nama tim kerja. Tabel ini terdiri dari Id\_Team serta Nama\_Team
- e. Tabel Fakta Rekap terdiri dari Id\_Pekerjaan, Id\_Petak, Id\_Waktu, Id\_team, Luas\_Lahan, dan Upah\_Satuan.

### 3) Data Preparation

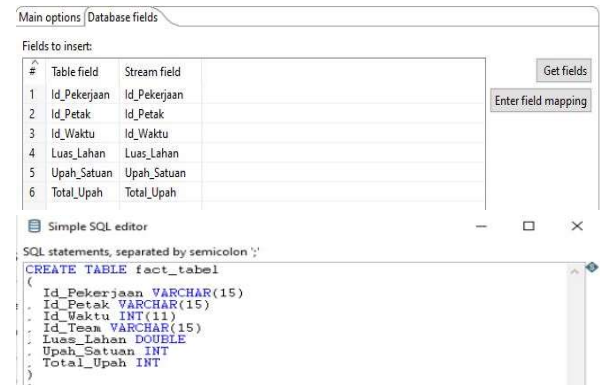
Fase ini merupakan tahapan yang digunakan dalam memperbaiki kesalahan-kesalahan yang ada pada data sehingga dapat menghasilkan *output* sesuai harapan. Dalam tahap ini, data hasil kerja tim akan yang bersumber dari Microsoft Excel akan dimasukkan ke dalam *database*. Tetapi sebelum data dimasukkan maka perlu dilakukan beberapa tahapan yang akan dijelaskan sebagai berikut:

- a. Proses Pembersihan data  
Proses ini digunakan untuk memperbaiki serta menghapus data-data yang terdapat kesalahan dan data yang tidak diperlukan dalam penelitian. Pada tahap ini dilakukan proses pembersihan data dari data yang tidak lengkap, data yang terdapat noisy serta menghapus data yang terdapat pengulangan.
- Tahap awal dilakukan pembersihan data menggunakan filter di Microsoft Excel untuk memfilter data-data yang tidak lengkap dan redundancy data. Pertama, dilakukan filter pada kolom Tanggal dan No\_SPK\_Manual lalu memunculkan kolom yang blanks untuk dihapus semua data yang tidak mempunyai nilai.
- Tahap selanjutnya dilakukan penghapusan nilai yang berulang menggunakan fitur conditional formatting pada Microsoft Excel dengan memilih fitur duplicate values sehingga akan muncul data-data berulang yang perlu dihapus.

Lalu dilakukan penyesuaian format tanggal dari bentuk yang lengkap dengan nama bulan menjadi bentuk sederhana yaitu dalam format dd/mm/yyyy dengan menggunakan fitur number format sehingga semua bentuk data tanggal menjadi sama.

Tahap akhir dilakukan transformasi data dari bentuk operasional menggunakan *tools* Pentaho Data Integration dengan memilih

*field-field* yang diperlukan lalu dilakukan *loading* data yang bersumber dari microsoft excel menjadi database MySQL dengan melakukan simple SQL editor pada *tools* Pentaho yang dapat dilihat pada Gambar 4.



Gambar 4. Proses Loading Data

Setelah melewati proses ETL (*Extraction, Transformation and Loadig*) maka didapatkan data yang telah bebas dari data *null* dan *redudancy* data. Data setelah pembersihan dapat dilihat pada Gambar 5.

A	B	C	D	E	F	G	H	I	J	K	L	M
No SPK	No SPK Manual	Jenis Pekerjaan	Id petak	PETAH	LOKASI	BAGAN	NAMA PT	TANGGAL	Luas	Upah Satuan	Total Upah	nama team
7126	51227694	500725	LANDU PUPUK MAMRE PER KG S-SAM	900725	BG0370	BAGAN NARE	PT BANI ANDALAS PERMANA	14/07/2020	41.150,0	Rp 66	Rp 2.716.900	MIAN
7127	51227694	500726	MIAT PUPUK + SUSUN KE DALAM TRUK	900726	BG0370	BAGAN NARE	PT BANI ANDALAS PERMANA	14/07/2020	41.150,0	Rp 30	Rp 1.234.500	MIAN
7128	51227694	500727	BONGKAR PULADAR TRUK + SUSUN DI GUDANG/TK/PT/TK	900727	BG0370	BAGAN NARE	PT BANI ANDALAS PERMANA	14/07/2020	41.150,0	Rp 30	Rp 1.234.500	MIAN
7129	51227694	500728	TRANSPORT BUBU POLYBAG AIR S-SAM	900728	BG0380	BAGAN NARE	PT BANI ANDALAS PERMANA	14/07/2020	32.980,0	Rp 13	Rp 428.830	MIAN
7130	51227694	500729	TRANSPORT BUBU POLYBAG DAMAT-H4-ASAM	900729	BG0380	BAGAN NARE	PT BANI ANDALAS PERMANA	14/07/2020	32.980,0	Rp 12	Rp 395.562	MIAN
7131	51227694	500730	BONGKAR BBT DAM KETEK + SUSUN DI PETAK	900730	BG0390	BAGAN NARE	PT BANI ANDALAS PERMANA	14/07/2020	32.980,0	Rp 5	Rp 168.900	MIAN
7132	51227694	500731	BONGKAR BBT DAM TRUK + SUSUN DI KETEK	900731	BG0390	BAGAN NARE	PT BANI ANDALAS PERMANA	14/07/2020	32.980,0	Rp 5	Rp 168.900	MIAN
7133	51227694	500732	LANDU PUPUK DAMAT PER KG H4-ASAM	900732	BG0390	BAGAN NARE	PT BANI ANDALAS PERMANA	14/07/2020	4.750,0	Rp 78	Rp 369.380	MIAN
7134	51227694	500733	LANDU PUPUK MAMRE PER KG S-SAM	900733	BG0390	BAGAN NARE	PT BANI ANDALAS PERMANA	14/07/2020	4.750,0	Rp 66	Rp 323.850	MIAN
7135	51227694	500734	MIAT PUPUK + SUSUN KE DALAM TRUK	900734	BG0390	BAGAN NARE	PT BANI ANDALAS PERMANA	14/07/2020	4.750,0	Rp 30	Rp 142.650	MIAN
7136	51227694	500735	BONGKAR PULADAR TRUK + SUSUN DI GUDANG/TK/PT/TK	900735	BG0390	BAGAN NARE	PT BANI ANDALAS PERMANA	14/07/2020	4.750,0	Rp 30	Rp 142.650	MIAN
7137	51227694	500736	LANDU PUPUK DAMAT PER KG H4-ASAM	900736	BG0390	BAGAN NARE	PT BANI ANDALAS PERMANA	14/07/2020	485,0	Rp 78	Rp 37.620	MIAN
7138	51227694	500737	LANDU PUPUK MAMRE PER KG S-SAM	900737	BG0390	BAGAN NARE	PT BANI ANDALAS PERMANA	14/07/2020	485,0	Rp 66	Rp 32.970	MIAN
7139	51227694	500738	MIAT PUPUK + SUSUN KE DALAM TRUK	900738	BG0390	BAGAN NARE	PT BANI ANDALAS PERMANA	14/07/2020	485,0	Rp 30	Rp 14.850	MIAN
7140	51227694	500739	BONGKAR PULADAR TRUK + SUSUN DI GUDANG/TK/PT/TK	900739	BG0390	BAGAN NARE	PT BANI ANDALAS PERMANA	14/07/2020	485,0	Rp 30	Rp 14.850	MIAN
7141	51227694	500740	LANDU PUPUK DAMAT PER KG H4-ASAM	900740	BG0390	BAGAN NARE	PT BANI ANDALAS PERMANA	14/07/2020	257,4	Rp 78	Rp 20.162	MIAN
7142	51227694	500741	LANDU PUPUK MAMRE PER KG S-SAM	900741	BG0390	BAGAN NARE	PT BANI ANDALAS PERMANA	14/07/2020	257,4	Rp 66	Rp 18.900	MIAN
7143	51227694	500742	MIAT PUPUK + SUSUN KE DALAM TRUK	900742	BG0390	BAGAN NARE	PT BANI ANDALAS PERMANA	14/07/2020	257,4	Rp 30	Rp 7.722	MIAN
7144	51227694	500743	BONGKAR PULADAR TRUK + SUSUN DI GUDANG/TK/PT/TK	900743	BG0390	BAGAN NARE	PT BANI ANDALAS PERMANA	14/07/2020	257,4	Rp 30	Rp 7.722	MIAN
7145	51227694	500744	TRANSPORT BUBU POLYBAG AIR S-SAM	900744	BG0470	BAGAN NARE	PT BANI ANDALAS PERMANA	14/07/2020	32.820,0	Rp 13	Rp 424.660	MIAN
7146	51227694	500745	TRANSPORT BUBU POLYBAG DAMAT-H4-ASAM	900745	BG0470	BAGAN NARE	PT BANI ANDALAS PERMANA	14/07/2020	32.820,0	Rp 12	Rp 395.240	MIAN
7147	51227694	500746	BONGKAR BBT DAM KETEK + SUSUN DI PETAK	900746	BG0470	BAGAN NARE	PT BANI ANDALAS PERMANA	14/07/2020	32.820,0	Rp 5	Rp 168.100	MIAN
7148	51227694	500747	BONGKAR BBT DAM TRUK + SUSUN DI KETEK	900747	BG0470	BAGAN NARE	PT BANI ANDALAS PERMANA	14/07/2020	32.820,0	Rp 5	Rp 168.100	MIAN
7149	51227694	500748	LANDU PUPUK DAMAT PER KG H4-ASAM	900748	BG0470	BAGAN NARE	PT BANI ANDALAS PERMANA	14/07/2020	6.244,9	Rp 78	Rp 478.422	MIAN
7150	51227694	500749	LANDU PUPUK MAMRE PER KG S-SAM	900749	BG0470	BAGAN NARE	PT BANI ANDALAS PERMANA	14/07/2020	6.244,9	Rp 66	Rp 427.180	MIAN
7151	51227694	500750	MIAT PUPUK + SUSUN KE DALAM TRUK	900750	BG0470	BAGAN NARE	PT BANI ANDALAS PERMANA	14/07/2020	6.244,9	Rp 30	Rp 239.340	MIAN

Gambar 5. Data Setelah Pembersihan

- Keterangan :
- a. Kriteria 1 menjelaskan mengenai “No SPK”
  - b. Kriteria 2 menjelaskan mengenai “Id Pekerjaan”
  - c. Kriteria 3 menjelaskan mengenai “Nama Pekerjaan”
  - d. Kriteria 4 menjelaskan mengenai “Nomor Petak”
  - e. Kriteria 5 menjelaskan mengenai “Nama Lokasi”
  - f. Kriteria 6 menjelaskan mengenai “Blok”
  - g. Kriteria 7 menjelaskan mengenai “Nama PT”
  - h. Kriteria 8 menjelaskan mengenai “Luas Lahan”
  - i. Kriteria 9 menjelaskan mengenai “Harga/Kegiatan”
  - j. Kriteria 10 menjelaskan mengenai “Nama Tim”
  - k. Kriteria 11 menjelaskan mengenai “Tanggal”

b. Proses Integrasi Data  
Pada tahap ini dilakukan penggabungan sumber data yang telah dibersihkan sehingga menjadi satu sumber data. Proses ini dilakukan dengan menggabungkan semua sheet file dari tahun 2019 sampai dengan 2022.

c. Proses Seleksi Data  
Setelah data dibersihkan dan digabungkan selanjutnya dilakukan proses seleksi data. Pada tahap ini dilakukan proses memilah data yang akan digunakan sesuai dengan kebutuhan

penelitian. Kolom yang dipilih adalah No\_SPK\_Sistem, Nama\_Pekerjaan, Tanggal, Lokasi, Nama\_Team, Luas\_Lahan serta Upah\_Satuan.

#### 4) Modelling

Tahap ini merupakan proses modelling dengan metode klasifikasi untuk menghasilkan suatu pola yang dapat memberikan informasi sehingga memudahkan pihak yang mempunyai kepentingan. Pola *Data Mining* ini digunakan untuk menghasilkan klasifikasi tim untuk ditempatkan di lokasi kerja tertentu. Pada tahap ini algoritma *Naive Bayes* menggunakan dapat diterapkan pada data yang berdifat kategorikal maupun nominal sehingga cocok untuk diterapkan pada data hasil kerja. Algoritma ini bekerja dengan cara setiap probabilitas pada variabel dianggap independen serta mengacu pada data historis setiap kelas. Algoritma ini juga hanya memerlukan sedikit data dalam menentukan estimasi parameter yang diperlukan untuk sistem klasifikasi yang kompleks sehingga lebih mudah dan efisien untuk diterapkan [20][21].

#### 5) Evaluation

Setelah didapatkan pola klasifikasi menggunakan algoritma *Naive Bayes* maka selanjutnya dilakukan evaluasi keakuratan penerapan algoritma *Naive Bayes* pada aplikasi yang telah dibuat. *Confusion matrix* digunakan untuk mengukur tingkat akurasi dari hasil klasifikasi yang terdiri dari: *Accuracy*, *precision*, dan *recall* dengan menggunakan persamaan (2), (3), dan (4) [22].

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

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

$$Accuracy = \frac{TP+TN}{TP+TN+FP+F} \quad (4)$$

#### 6) Deployment

Tahap deployment adalah dihasilkan tahap menilai hasil dari model *Naive Bayes* yang telah diterapkan lalu dilakukan penyesuaian model sehingga menghasilkan output sesuai dengan target dan diharapkan dapat mendukung penunjang keputusan yang dapat digunakan oleh pihak perusahaan untuk memberikan pertimbangan dalam penentuan langkah terkait upaya menempatkan tim kerja dilokasi sehingga dapat meningkatkan kualitas dari perusahaan.

### 3. HASIL

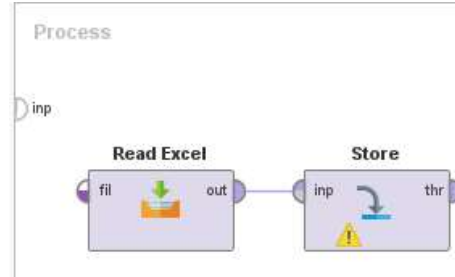
Penelitian ini menggunakan data hasil kerja dengan total 118.154 data. Sebelum dilakukan metode *Naive Bayes* dalam klasifikasi tim kerja maka perlu dilakukan tahap selection fitur, integrasi data serta transformasi data. Pada tahap selection fitur dilakukan penghapusan atribut-atribut yang tidak diperlukan seperti bulan kerja, total upah dan nomor spk. Lalu tahap integrasi data dilakukan dengan memilih atribut yang digunakan sebagai *primary key*. Lalu transformasi data dilakukan dengan mengubah data dari bentuk operasional mircosoft excel menjadi bentuk *database*.

#### 3.1.1. Model Pengujian Pada Naive Bayes

Pada tahap ini dilakukan pengujian untuk mengetahui apakah

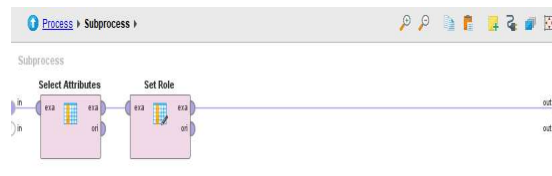
data hasil kerja yang digunakan dapat dilakukan pengujian menggunakan metode klasifikasi dengan algoritma *Naive Bayes*. Pengujian dilakukan dengan menentukan *data testing* yang telah dipilih. Dalam mencari nilai prediksi data testing menggunakan *tools* Rapid Miner. Lalu dilakukan pula pemilihan *attributes* untuk mengetahui hasil analisa dari Rapid Miner.

Gambar 6 merupakan tampilan pada saat melakukan pengujian algoritma *Naive Bayes* dengan *tools* Rapidminer, langkah awal menggunakan operaotor Read Excel yang digunakan untuk memanggil data hasil kerja. Dataset dibagi menjadi *data training* dan *data testing*. *Data training* digunakan untuk melatih model *Naive Bayes* sedangkan *data testing* digunakan untuk menguji model *Naive Bayes*.



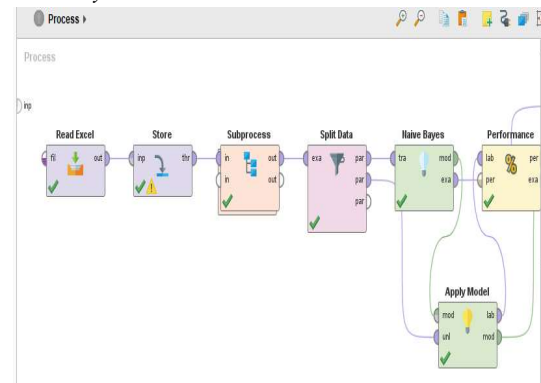
Gambar 6. Memanggil Data

Pada Gambar 7 dilakukan pemilihan atribut data yang akan digunakan lalu dilakukan *set role* untuk membedakan antara kolom yang akan di gunakan sebagai label dari data hasil kerja. Kolom yang digunakan sebagai label adalah kolom nama tim.



Gambar 7. Subprocess Detail

Gambar 8 menunjukkan pengujian dengan menggunakan algoritma *Naive Bayes* pada Rapidminer. Selanjutnya digunakan operator Performance untuk mengeluarkan hasil dari kinerja model *Naive Bayes*. Dan terdapat operator *Apply Model* yang digunakan untuk mengeluarkan rule model *Naive Bayes*.

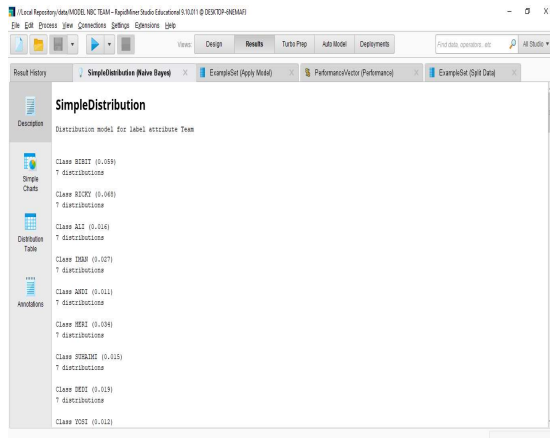


Gambar 8. Model Naive Bayes

Gambar 9 menunjukkan hasil *simple distribution* setiap class-class pada label nama tim. *Simple distribution* digunakan untuk melakukan analisis terhadap data hasil kerja



dalam memprediksi nama tim.



Gambar 9 Hasil Simple Distribution

Gambar 10 merupakan hasil pengujian menggunakan tool rapid miner yang menunjukkan hasil dari model klasifikasi algoritma *Naive Bayes*.

The table is a confusion matrix with 'accuracy 93.66%' at the top. The columns represent predicted classes (pred.BIB, pred.PIC, pred.ALI, pred.MAN, pred.ANCI, pred.HERI, pred.SLAH, pred.DEDI, pred.YOSB, pred.PADI, pred.KUSI, pred.KAN) and the rows represent actual classes (act.BIB, act.PIC, act.ALI, act.MAN, act.ANCI, act.HERI, act.SLAH, act.DEDI, act.YOSB, act.PADI, act.KUSI, act.KAN). The diagonal elements are highlighted in yellow, indicating correct classifications. The overall accuracy is 93.66%.

Gambar 10 Hasil Pengujian

Untuk mengetahui tingkat akurasi dari data yang dibangun maka digunakan informasi dari *Confusion matrix* yang dapat dilihat pada tabel 1.

Tabel 2. Hasil Evaluasi

Evaluasi	Persentase
Accuracy	93,66%
Recall	90,74%
Precision	88,58%

Dari tabel 2 dapat dilihat hasil evaluasi dengan Nilai accuracy yang dihasilkan dari perhitungan seluruh nilai prediksi yang benar dibagi jumlah seluruh data. Nilai accuracy yang dihasilkan cukup tinggi yaitu sebesar 93,66%. Lalu hasil recall yang dihasilkan dari jumlah prediksi positif benar dibagi dengan jumlah seluruh kelas positif yaitu sebesar 90,74%. Selanjutnya nilai precision dihasilkan dari seluruh jumlah nilai prediksi positif benar dibagi seluruh prediksi kelas benar yang menghasilkan nilai sebesar 88,58%.

#### 4. PEMBAHASAN

##### 4.1. Implementasi Data Mining

Pada tahap ini dilakukan implementasi aplikasi *Data Mining* berbasis website. Pada halaman dataset menunjukkan data-data yang ada pada *database MySQL*. Pada halaman ini berisi atribut-atribut seperti jenis pekerjaan, petak, lokasi, nama PT, luas, Upah satuan dan nama tim. Sample data berisi 118.154 baris data yang dapat dicari dengan bantuan search. Halaman sample dataset dapat dilihat pada gambar 11.

The screenshot shows a 'Sample Dataset' table with 7 columns: 'Jenis Pekerjaan', 'Petak', 'Lokasi', 'Nama PT', 'Luas', 'Upah Satuan', and 'Nama Tim'. The data rows show entries for 'BLANKING (SULAM)' with various 'Petak' numbers (e.g., ASH2300, ASH2140, ASH4100) and 'Upah Satuan' values (e.g., 11.7, 22.8, 22.9). The last row is 'CHEMICAL WEEDING 1' with 'Petak' ASH1190 and 'Upah Satuan' Rp197.700.

Gambar 11. Halaman Dataset

Gambar 12 menunjukkan halaman tambah data baru yang digunakan untuk menambahkan data jenis pekerjaan baru, lokasi baru dan nama tim baru yang akan bertambah di kemudian hari sehingga aplikasi ini dapat dipakai hingga tahun-tahun berikutnya.

## Tambah Data Baru

Jenis Pekerjaan

Lokasi

Nama Tim

**Tambah**

Gambar 12. Halaman Tambah Data Baru

Gambar 13 menunjukkan halaman klasifikasi menggunakan algoritma *Naive Bayes* yang digunakan untuk merekomendasikan tim yang sebaiknya dipilih untuk melakukan proyek pekerjaan selanjutnya. Pada halaman ini berisi :

- 1) Jenis Pekerjaan dalam bentuk combo box yang telah berisi semua pekerjaan yang sudah pernah dilakukan.

Perusahaan dapat memilih jenis pekerjaan yang akan dilaksanakan selanjutnya.

- 2) Lokasi dalam bentuk combo box yang berisi nama-nama lokasi kerja yang terdapat pada PT.MNL Perusahaan dapat memilih lokasi kerja proyek selanjutnya.
- 3) Lalu terdapat button submit yang digunakan untuk menghitung algoritma *Naive Bayes* yang akan ditampilkan pada rekomendasi tim.
- 4) Terdapat perhitungan *Naive Bayes* dengan memunculkan nama lima team dengan hasil tertinggi.

Setelah muncul rekomendasi tim, perusahaan dapat tidak memilih tim tersebut karena sifatnya hanya sebuah rekomendasi untuk perusahaan sehingga nama tim dapat dipilih lagi di dalam *combobox* pilih tim kerja. Selanjutnya terdapat *button* simpan hasil, data yang disimpan akan masuk ke dalam *database* MySQL.

Penentuan Tim Kerja

Jenis Pekerjaan

Lokasi

**SUBMIT**

Perhitungan Naive Bayes

Tim	P(Tim)
BIBIT	7015/ 118062 = 0.059417932950484
RECKY	8060/ 118062 = 0.068269214480527
ALI	1919/ 118062 = 0.016254171536989
IMAN	3149/ 118062 = 0.026672426352256
ANDI	1297/ 118062 = 0.010985753248293

P(JenisPekerjaan Tim)	BIBIT	RECKY	ALI	IMAN	ANDI
SINGLING + TEER	208/7015 = 0.029650748396294	302/8060 = 0.037468982630273	54/1919 = 0.02813965607087	93/3149 = 0.029533185138139	30/1297 = 0.0231303
MANJUAL WEEDING 1	155/7015 = 0.02209509622238	175/8060 = 0.021712158808933	36/1919 = 0.018759770713913	4/3149 = 0.0012702445220705	0/1297 = 0
MANJUAL WEEDING 2	67/7015 = 0.00959509622238061	7/8060 = 0.00086848639235732	19/1919 = 0.009900990990099	0/3149 = 0	0/1297 = 0
MANJUAL WEEDING 3	21/7015 = 0.0029939851746258	0/8060 = 0	11/1919 = 0.0057321521625847	0/3149 = 0	0/1297 = 0
TRANSPORT BIBIT POLYTUBE AIR ID-19.99 KM	18/7015 = 0.0025659301496793	28/8060 = 0.0034739454094293	2/1919 = 0.0010422094841063	0/3149 = 0	0/1297 = 0

P(Lokasi Tim)	BIBIT	RECKY	ALI	IMAN	ANDI
SINGLING + TEER	1146/7015 = 0.16336421952958	4377/8060 = 0.54305210918114	0/1919 = 0	1/3149 = 0.00031756113051762	0/1297 = 0
MANJUAL WEEDING 1	1404/7015 = 0.20014295167498	0/8060 = 0	1090/1919 = 0.56800416883794	1192/3149 = 0.37853286757701	792/1297 = 0.57979953739399
MANJUAL WEEDING 2	1632/7015 = 0.23264433357092	1665/8060 = 0.20657568238213	0/1919 = 0	862/3149 = 0.27373769450619	0/1297 = 0
MANJUAL WEEDING 3	2132/7015 = 0.30392017106201	0/8060 = 0	0/1919 = 0	684/3149 = 0.21721181327406	378/1297 = 0.29144178874325
TRANSPORT BIBIT POLYTUBE AIR ID-19.99 KM	673/7015 = 0.095937277263008	2018/8060 = 0.25037220843672	0/1919 = 0	204/3149 = 0.064782470625595	0/1297 = 0

$P(\text{new}|\text{BIBIT}) = 0.00021447590272132$   
 $P(\text{new}|\text{RECKY}) = 0.00080495095040487$   
 $P(\text{new}|\text{ALI}) = 0$   
 $P(\text{new}|\text{IMAN}) = 1.0759130982624E-8$   
 $P(\text{new}|\text{ANDI}) = 0$   
 Rekomendasi Tim : Tergolong Pada Tim Risky

Nama Tim

**STAPAN HASIL**

Gambar 13. Halaman Klasifikasi Tim

## 5. KESIMPULAN

Implementasi *Data Mining* pada metode klasifikasi dengan menggunakan algoritma *Naive Bayes* dapat menemukan pola dari data yang memiliki karakteristik yang sama. Data yang digunakan adalah data hasil kerja PT. MNL. Dari tabel evaluasi dihasilkan nilai *accuracy* sebesar 93,66%, nilai *recall* 90,74% serta nilai *precision* sebesar 88,58%. Aplikasi *data mining* yang dirancang dengan menggunakan pemrograman PHP dan *database* MySQL mampu menampilkan informasi dalam mengklasifikasi tim kerja dengan menggunakan data hasil kerja yang ada sebagai *data training* dan pekerjaan baru sebagai *data testing* sehingga dapat digunakan oleh pemilik perusahaan sebagai bahan pertimbangan dalam mengatur persebaran tim di setiap lokasi kerja.

## DAFTAR PUSTAKA

- [1] H. T. Sherly, Yosef Stephen, "Analisis dan Perancangan Data Warehouse Pada Koperasi IPEKA," *Encyclopedia of Geographic Information Science*, pp. 1–6, 2018, doi: [10.4135/9781412953962.n41](https://doi.org/10.4135/9781412953962.n41).
- [2] H. Lubis and D. B. Srisulistiwati, "Implementasi Sistem Data Mining Untuk Menentukan Stock Accuracy Pada Warehouse PT Coca-Cola Amatil Indonesia Cibitung-Plant," *Jurnal Media Informatika Budidarma*, vol. 4, no. 1, p. 72, 2020, doi: [10.30865/mib.v4i1.1795](https://doi.org/10.30865/mib.v4i1.1795).
- [3] M. Kusmawaty, "Perancangan dan implementasi data mining dalam proses penerimaan siswa baru dengan metode quantitative association rule," *Jurnal Teknik Komputer*, 2017.
- [4] Retroningsih, "Bab II Landasan Teori," *J Chem Inf Model*, vol. 53, no. 9, pp. 8–24, 2019.
- [5] D. Puspitasari, S. S. Al Khautsar, and W. P. Mustika, "Algoritma Naive Bayes Untuk Memprediksi Kredit Macet Pada Koperasi Simpan Pinjam," *Jurnal Informatika Upgris*, vol. 4, no. 2, 2019, doi: [10.26877/jiu.v4i2.2919](https://doi.org/10.26877/jiu.v4i2.2919).
- [6] A. W. Syaputri and E. Irwandi, "Naive Bayes Algorithm for Classification of Student Major's Specialization," *Journal of Intelligent Computing and Health Informatics*, vol. 1, no. 1, pp. 1–5, 2020.
- [7] A. Saleh, "Implementasi Naive Bayes," *Journal of Informatics, Information System, Software Engineering and Applications (INISTA)*, vol. 2, no. 3, pp. 73–81, 2015, doi: [10.20895/inista.v1i2.73](https://doi.org/10.20895/inista.v1i2.73).
- [8] E. Manalu, F. A. Sianturi, and M. R. Manalu, "Penerapan Algoritma Naive Bayes Untuk Memprediksi Jumlah Produksi Barang Berdasarkan Data Persediaan Dan Jumlah Pemesanan Pada CV. Papadan Mama Pastries," *Jurnal Mantik Penusa*, vol. 1, no. 2, pp. 16–21, 2017.
- [9] P. Han, Kamber, *Data mining: Data mining concepts and techniques*. 2014. doi: [10.1109/ICMIRA.2013.45](https://doi.org/10.1109/ICMIRA.2013.45).
- [10] H. F. Putro, R. T. Vlandari, and W. L. Saptomo, "Penerapan Metode Naive Bayes Untuk Klasifikasi Pelanggan," vol. 8, no. 2, 2020.
- [11] F. Fitriyani, "Implementasi Forward Selection dan Bagging untuk Prediksi Kebakaran Hutan Menggunakan Algoritma Naive Bayes," *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 8, no. 1, pp. 1–8, 2022, doi: [10.25077/teknosi.v8i1.2022.1-8](https://doi.org/10.25077/teknosi.v8i1.2022.1-8).
- [12] E. Y. Hidayat and M. A. Rizqi, "Klasifikasi Dokumen Berita Menggunakan Algoritma Enhanced Confix Stripping Stemmer dan Naive Bayes Classifier," *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 6, no. 2, pp. 90–99, 2020, doi: [10.25077/teknosi.v6i2.2020.90-99](https://doi.org/10.25077/teknosi.v6i2.2020.90-99).
- [13] M. Silvana, R. Akbar, and A. Syahnum, "Pemanfaatan Metode Naive Bayes dalam Implementasi Sistem Pakar Untuk Menganalisis Gangguan Perkembangan Anak," *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 6, no. 2, pp. 74–81, 2020, doi: [10.25077/teknosi.v6i2.2020.74-81](https://doi.org/10.25077/teknosi.v6i2.2020.74-81).
- [14] Abiyoga, A. Wicaksana, and N. M. S. Iswari, "Decision Support System for Choosing an Elective Course Using Naive Bayes Classifier," in *Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing*, R. Lee, Ed., Cham: Springer International Publishing, 2020, pp. 97–110. doi: [10.1007/978-3-030-26428-4\\_7](https://doi.org/10.1007/978-3-030-26428-4_7).
- [15] Moh. S. Anam, "Social Assistance Decision Support System Using the Naive Bayes Method," *JTECS: Jurnal Sistem Telekomunikasi Elektronika Sistem Kontrol Power Sistem dan Komputer*, vol. 1, no. 1, p. 85, 2021, doi: [10.32503/jtecs.v1i1.1433](https://doi.org/10.32503/jtecs.v1i1.1433).
- [16] D. M. Khairina, S. Maharani, Ramadiani, and H. R. Hatta, "Decision support system for admission selection and positioning human resources by using naive bayes method," *Adv Sci Lett*, vol. 23, no. 3, pp. 2495–2497, 2017, doi: [10.1166/asl.2017.8653](https://doi.org/10.1166/asl.2017.8653).
- [17] F. Burdi, A. H. Setianingrum, and N. Hakiem, "Application of the naive bayes method to a decision support system to provide discounts (Case study: PT. Bina Usaha Teknik)," *Proceedings - 6th International Conference on Information and Communication Technology for the Muslim World, ICT4M 2016*, pp. 281–285, 2017, doi: [10.1109/ICT4M.2016.57](https://doi.org/10.1109/ICT4M.2016.57).
- [18] A. P. Fadillah, "Penerapan Metode CRISP-DM untuk Prediksi Kelulusan Studi Mahasiswa Menempuh Mata Kuliah (Studi Kasus Universitas XYZ)," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 1, no. 3, pp. 260–270, 2015, doi: [10.28932/jutisi.v1i3.406](https://doi.org/10.28932/jutisi.v1i3.406).
- [19] M. A. Hasanah, S. Soim, and A. S. Handayani, "Implementasi CRISP-DM Model Menggunakan Metode Decision Tree dengan Algoritma CART untuk Prediksi Curah Hujan Berpotensi Banjir," *Journal of Applied Informatics and Computing*, vol. 5, no. 2, pp. 103–108, 2021.
- [20] M. H. Adiya and Y. Desnelita, "Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan Pada RSUD Pekanbaru," vol. 01, pp. 17–24, 2019.
- [21] M. K. Suryadewiansyah and T. E. E. Tju, "Naive Bayes dan Confusion Matrix untuk Efisiensi Analisa Intrusion Detection System Alert," *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 8, no. 2, pp. 81–88, 2022, doi: [10.25077/teknosi.v8i2.2022.81-88](https://doi.org/10.25077/teknosi.v8i2.2022.81-88).
- [22] D. K. Sharma, M. Chatterjee, G. Kaur, and S. Vavilala, "3 - Deep learning applications for disease diagnosis," D. Gupta, U. Kose, A. Khanna, and V. E. B. T.-D. L. for



M. A. with U. D. Balas, Eds., Academic Press, 2022, pp. 31–51. doi: [10.1016/B978-0-12-824145-5.00005-8](https://doi.org/10.1016/B978-0-12-824145-5.00005-8).

## BIODATA PENULIS



Tasya Angelya

Saat ini penulis sebagai mahasiswa tahap akhir di Universitas Multi Data Palembang, Fakultas Ilmu Komputer dan Rekayasa, Program Studi Sistem Informasi. Ketertarikan dalam bidang data membuat penulis melakukan penelitian terkait *data mining* dengan metode klasifikasi.



Abdul Rahman, S.Si, M.T.I

Penulis menyelesaikan pendidikan sarjana di Program Studi Elektronika dan Instrumentasi Fakultas MIPA, Universitas Gadjah Mada sebagai Sarjana Sains (S.Si). Kemudian melanjutkan studi ke jenjang S2 Magister Teknologi Informasi di Universitas Indonesia dengan bidang data warehouse dan data mining



Iis Pradesan, S.Kom, M.T.I

Penulis menyelesaikan pendidikan sarjana di Program Studi Sistem Informasi STMIK Multi Data Palembang sebagai Sarjana Komputer (S.Kom). Kemudian melanjutkan studi ke jenjang S2 Magister Teknologi Informasi di Universitas Indonesia dengan bidang Teknologi Informasi