

**PENERAPAN ALGORITMA NAIVE BAYES DALAM
MEMPREDIKSI PENJUALAN MATERIAL
BANGUNAN**

SKRIPSI



**Oleh:
Sandy
170210056**

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK DAN KOMPUTER
UNIVERSITAS PUTERA BATAM
TAHUN 2022**

**PENERAPAN ALGORITMA NAIVE BAYES DALAM
MEMPREDIKSI PENJUALAN MATERIAL
BANGUNAN**

SKRIPSI

**Untuk memenuhi salah satu syarat
Memperoleh gelar Sarjana**



**Oleh:
Sandy
170210056**

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK DAN KOMPUTER
UNIVERSITAS PUTERA BATAM
TAHUN 2022**

SURAT PERNYATAAN ORISINALITAS

Yang bertanda tangan di bawah ini saya

Nama : Sandy
NPM : 170210056
Fakultas : Teknik dan Komputer
Program Studi : Teknik Informatika

Menyatakan bahwa “Skripsi” yang saya buat dengan judul:

PENERAPAN ALGORITMA NAIVE BAYES DALAM MEMPREDIKSI PENJUALAN MATERIAL BANGUNAN

Adalah hasil karya sendiri dan bukan “duplikasi” dari karya orang lain. Sepengantahuan saya, di dalam naskah Skripsi ini tidak terdapat karya ilmiah atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis dikutip didalam naskah ini dan disebutkan dalam sumber kutipan dan daftar pustaka.

Apabila ternyata di dalam naskah Skripsi ini dapat dibuktikan terdapat unsur-unsur PLAGIASI, saya bersedia naskah Skripsi ini digugurkan dan gelar akademik yang saya peroleh dibatalkan, serta proses sesuai dengan peraturan perundang-undangan yang berlaku.

Demikian pernyataan ini saya buat dengan sebenarnya tanpa ada paksaan dari siapapun.

Batam, 01 Agustus 2022



Sandy
170210056

**PENERAPAN ALGORITMA NAIVE BAYES DALAM MEMPREDIKSI
PENJUALAN MATERIAL BANGUNAN**

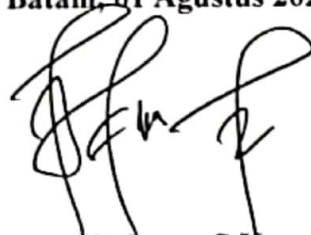
SKRIPSI

**Untuk memenuhi salah satu syarat
memperoleh gelar Sarjana**

**Oleh :
Sandy
170210056**

**Telah disetujui oleh Pembimbing pada tanggal
seperti tertera di bawah ini**

Batam, 01 Agustus 2022



**Sunarsan Sitohang, S.Kom., M.TI.
Pembimbing**

ABSTRAK

Material merupakan bahan bangunan berupa benda atau objek yang digunakan untuk membangun tempat tinggal seperti suatu rumah yang digunakan sebagai tempat tinggal dan tempat perlindungan dari musim hujan, panas dan sebagainya. Setiap proses pembangunan rumah perlu adanya bahan bangunan. Siklus penjualan bahan bangunan yaitu *distributor* menyediakan bahan material yang *disupplier* ke toko bangunan dan konsumen membeli bahan material di toko bangunan terdekat. Setiap perusahaan memiliki penyimpanan data dan data yang disimpan semakin lama berjalan semakin menumpuknya datanya. Data penyimpanan akan dimanfaatkan untuk menerapkan data *mining*. Tujuan dari penelitian ini adalah untuk mengetahui hasil penerapan data *mining* di PT.Tanjung Uncang dalam memprediksi penjualan bahan material yang laku dan tidak laku menggunakan metode *Naïve Bayes*. Dalam perancangan desain penelitian memerlukan sebuah desain sebagai arah penjas, dimulai dengan cara memperoleh data, mengolah data, dan langkah-langkah untuk menyelesaikan pengolahan data. Data penjualan pada PT.Tanjung Uncang menggunakan metode *naïve bayes* serta di uji melalui aplikasi *WEKA 3.9* diperoleh hasil yaitu 135 yang laku sebanyak 72.5806% dan 51 yang tidak laku sebanyak 27.4194% dari keseluruhan 186 data dengan 12 variabel yaitu variable Januari, Februari, Maret, April, Mei, Juni, Juli, Agustus, September, Oktober, November, dan Desember. Dengan menggunakan teknik data *mining* dan *algoritma Naïve Bayes* dapat mempercepat pengolahan data penjualan sehingga memiliki peran penting dalam membantu memdapatkan informasi laku dan tidak laku dimasa yang akan datang.

Kata Kunci: *Data Mining*, Aplikasi *WEKA*, *Naïve Bayes*.

ABSTRACT

Materials are building materials in the form of objects or objects that are used to build residences such as a house that is used as a residence and a place of protection from the rainy season, heat and so on. Every process of building a house requires building materials. The sales cycle of building materials is that distributors provide materials that are supplied to building stores and consumers buy materials at the nearest building store. Every company has data storage and the data that is stored the longer the data accumulates. Storage data will be used to implement data mining. The purpose of this study was to determine the results of the application of data mining at PT. Tanjung Uncang in predicting sales of materials that were and did not sell using the Naïve Bayes method. In designing the research design requires a design as an explanatory direction, starting with obtaining data, processing data, and steps to complete data processing. Sales data at PT. Tanjung Uncang using the nave Bayes method and tested through the WEKA 3.9 application, the results obtained are 135 which are sold as much as 72,5806% and 51 are not sold as much as 27.4194% of a total of 186 data with 12 variables, namely January, February, March, April, May, June, July, August, September, October, November, and December. By using data mining techniques and yahoo Nave Bayes can accelerate the processing of sales data so that it has an important role in helping to get information about behavior and not selling in the future.

Keyword: *Data Mining, WEKA Application, Naive Bayes*

KATA PENGANTAR

Puji syukur kehadiran Tuhan Yang Maha Esa yang telah melimpahkan segala rahmat dan karuniaNya, sehingga penulis dapat menyelesaikan laporan tugas akhir yang merupakan salah satu persyaratan untuk menyelesaikan program studi strata satu (S1) pada Program Studi Teknik Informatika Universitas Putera Batam.

Penulis menyadari bahwa skripsi ini masih jauh dari sempurna. Karena itu, kritik dan saran akan senantiasa penulis terima dengan senang hati. Dengan segala keterbatasan, penulis menyadari pula bahwa skripsi ini takkan terwujud tanpa bantuan, bimbingan, dan dorongan dari berbagai pihak. Untuk itu, dengan segala kerendahan hati, penulis menyampaikan ucapan terima kasih kepada:

1. Rektor Universitas Putera Batam, Ibu Nur Elfi Husda, S.Kom., M.SI.;
2. Dekan Fakultas Teknik dan Komputer, Bapak Welly Sugianto, S.T., M.M.;
3. Ketua Program Studi Teknik Informatika, Bapak Andi Maslan, S.T., M.SI.;
4. Bapak Sunarsan Sitohang, S.Kom., M.TI. selaku pembimbing Skripsi pada Program Studi Teknik Informatika Universitas Putera Batam;
5. Bapak Koko Handoko, S.Kom., M.Kom. selaku pembimbing Akademik pada Program Studi Teknik Informatika Universitas Putera Batam;
6. Dosen dan Staff Universitas Putera Batam;
7. Semua anggota keluarga yang telah banyak memberikan dukungan dan kontribusi kepada penulis;
8. Semua rekan-rekan kelas IT Nagoya angkatan 2017 yang memberikan semangat dan bantuan selama penyusunan laporan ini;
9. Serta semua pihak yang baik secara langsung maupun tidak langsung yang telah membantu penulisan dalam penyusunan laporan ini.

Semoga Tuhan Yang Maha Esa membalas kebaikan dan selalu mencurahkan hidayah serta taufik-Nya, Amin.

Batam, 01 Agustus 2022



Sandy

DAFTAR ISI

	Halaman
HALAMAN SAMPUL	i
HALAMAN JUDUL	ii
SURAT PERNYATAAN ORISINALITAS	iii
HALAMAN PENGESAHAN	iv
ABSTRAK	v
ABSTRACT	vi
KATA PENGANTAR	vii
DAFTAR ISI	viii
DAFTAR GAMBAR	x
DAFTAR TABEL	xi
DAFTAR RUMUS	xii
BAB I PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Identifikasi Masalah	3
1.3. Batasan Masalah.....	4
1.4. Rumusan Masalah	4
1.5. Tujuan Penelitian.....	5
1.6. Manfaat Penelitian.....	5
1.6.1. Manfaat Teoritis.....	5
1.6.2. Manfaat Praktis	5
BAB II KAJIAN PUSTAKA	7
2.1. Konsep Teoritis	7
2.1.1. <i>Knowledge Discovery in Database (KDD)</i>	7
2.2. <i>Data Mining</i>	9
2.2.1. Pengenalan <i>Data Mining</i>	9
2.2.2. Proses <i>Data Mining</i>	9
2.3. Metode <i>Data Mining</i>	11
2.4. Software Pendukung	12
2.5. Penelitian Terdahulu	13
2.6. Kerangka Berfikir.....	18

BAB III METODE PENELITIAN	19
3.1 Desain Penelitian	19
3.2 Teknik Pengumpulan Data	20
3.3 Operasional Variabel	21
3.4 Metode Perancangan Sistem	22
3.5 Lokasi dan Jadwal Penelitian	23
BAB IV HASIL DAN PEMBAHASAN	23
4.1 Hasil Penelitian	25
4.2 Arsitektur Sistem dan Pembahasan	26
4.2.1 Seleksi Data dan Eliminasi Atribut	33
4.2.2 Pengujian data (<i>Data Testing</i>)	37
4.3 Pengujian Data Menggunakan Aplikasi <i>WEKA</i> 3.9	55
BAB V KESIMPULAN DAN SARAN	62
5.1 Kesimpulan	62
5.2 Saran	63
DAFTAR PUSTAKA	64
LAMPIRAN	66
Lampiran 1. Pendukung Penelitian	66
Lampiran 2. Daftar Riwayat Hidup	76
Lampiran 3. Surat Keterangan Penelitian	77
Lampiran 4. Hasil Turnitin Skripsi	79
Lampiran 5. Jurnal Dan Hasil Turnitin Jurnal	80

DAFTAR GAMBAR

	Halaman
Gambar 2.1 Knowladge Discovery in Database (KDD)	7
Gambar 2.2 Kerangka Berfikir	18
Gambar 3.1 Desain Penelitian	19
Gambar 3.2 Metode Perancangan Sistem.....	22
Gambar 3.3 Lokasi PT.Tanjung Uncang	24
Gambar 4.1 Tampilan Excel disimpan ke CSV.....	56
Gambar 4. 2 Tampilan Data Kedalam Notepad++.....	56
Gambar 4.3 Tampilan Data Convert CSV to ARFF	57
Gambar 4.4 Tampilan Aplikasi WEKA 3.9	57
Gambar 4.5 Membuka Data Pada Menu Explorer	58
Gambar 4.6 Pemilihan Atribut Pada Aplikasi WEKA.....	58
Gambar 4.7 Pemilihan Model Classify	59
Gambar 4.8 Pemilihan Naïve Bayes.....	59
Gambar 4.9 Pemilihan Tombol Start.....	60
Gambar 4.10 Hasil Pengujian Dengan Aplikasi WEKA 3.9.....	61

DAFTAR TABEL

	Halaman
Tabel 3.1 Jadwal Penelitian.....	24
Tabel 4.1 Data Penjualan Tahun 2019	27
Tabel 4.2 Data Penjualan Tahun 2020	28
Tabel 4.3 Data Penjualan Tahun 2021	29
Tabel 4.4 Data Kategori Laku Dan Tidak Laku Sementara	31
Tabel 4.5 Hasil Data Kategori Laku dan Tidak Laku Sementara.....	33
Tabel 4.6 Data Training Penjualan Tahun 2019 Setelah Eliminasi	34
Tabel 4.7 Data Training Penjualan Tahun 2020 Setelah Eliminasi	35
Tabel 4.8 Data Training Penjualan Tahun 2021 Setelah Eliminasi	36
Tabel 4.9 Data Testing	37
Tabel 4.10 Data testing 1.....	38
Tabel 4.11 Hasil Data testing 1	41
Tabel 4.12 Data Testing 2	41
Tabel 4.13 Hasil Data Testing 2.....	44
Tabel 4.14 Data Testing 3	45
Tabel 4.15 Hasil Data Testing 3.....	48
Tabel 4.16 Data Testing 4	48
Tabel 4.17 Hasil Data Testing 4.....	51
Tabel 4.18 Data Testing 5	52
Tabel 4.19 Hasil Data Testing 5.....	55

DAFTAR RUMUS

	Halaman
Rumus 2.1 Probability Bayes	11
Rumus 2.2 Teorema Bayes	11
Rumus 2.3 <i>Probability Bayes</i>	12
Rumus 2.4 <i>Teorema Bayes</i>	12
Rumus 4.1 Jumlah Barang	30
Rumus 4.2 Rata-Rata	30
Rumus 4.3 Laku dan Tidak Laku	30

BAB I

PENDAHULUAN

1.1. Latar Belakang

Material merupakan bahan-bahan bangunan berupa benda yang digunakan untuk membangun tempat tinggal seperti suatu rumah digunakan sebagai tempat perlindungan dari musim hujan, panas dan sebagainya. Di dalam bidang material, Setiap proses pembangunan rumah perlu adanya bahan bangunan yang bisa dicari ditoko bangunan terdekat, contoh bahan bangunan seperti papan gypsum, baja ringan, spandek, paku dan sebagai berikut. Siklus penjualan bangunan yaitu distributor menyediakan bahan material yang disupplier ke toko bangunan dan konsumen membeli bahan material di toko bangunan. setiap tahunnya itu terjadi peningkatan, karena banyak pembangunan proyek, renovasi rumah dan terutama di perkotaan seperti kota di Batam. Tidak diherankan apabila *distributor* penjualan bahan material dan toko bangunan yang di Batam semakin meningkat.

Di era modern sekarang, kebanyakan perusahaan memanfaatkan teknologi mesin yang sudah dirancang untuk mencetak bahan bangunan dengan menggunakan bahan baku, seperti bahan baku yang telah tersedia di alam berupa pohon yang diolah menjadi papan. Lalu dari segi teknologi informasi adalah untuk mempermudah pengusaha dalam pengolahan data. Setiap perusahaan pasti memiliki penyimpanan data dan data yang disimpan semakin lama berjalan maka semakin menumpuknya datanya. Oleh itu `peneliti akan memanfaatkan data-data

yang telah disimpan agar bisa digunakan untuk memprediksi penjualan bahan material.

Perseroan terbatas (PT) Tanjung Uncang merupakan salah satu *distributor* yang menyediakan bahan material ke toko-toko bangunan dan adapun kompetitor yang bergerak dibidang yang sama sehingga penjualan sering terjadi perbandingan harga, kualitas dan lainnya. Disetiap hari pengiriman barang PT Tanjung Uncang memenuhi kebutuhan konsumen, akan tetapi ada permasalahan yang terjadi yaitu barang yang tidak laku tidak bisa dijual sehingga menjadi stock mati dan lama kelamaan barang tersebut bisa rusak. Maka dari itu harus segera menjualnya atau tidak perlu memasukan barang yang tidak laku lagi. Dipenjualan barang material PT Tanjung Uncang hanya disimpan datanya sehingga akan terjadi penumpukan data dan data penjualan tersebut tidak dimanfaatkan ataupun menerapkan data *mining*. Untuk dapat mengatasi masalah tersebut, peneliti akan memanfaatkan data-data penjualan yang tersimpan di PT Tanjung Uncang. Data-data penjualan PT Tanjung Uncang akan dijadikan dasar memprediksi penjualan material yang akan datang dengan menerapkan data *mining* yang memakai algoritma *Naive Bayes*.

Penerapan data mining adalah proses pendataan dalam jumlah yang sangat besar dalam memperkerjakan suatu database untuk membantu menganalisis data penjualan dalam menentukan hubungan atau pola yang bermanfaat untuk pengusaha. (Listriani et al., 2016). Dengan data *mining*, penjualan pada bahan material dapat dioptimalkan melalui pendataan.

Dalam penulisan skripsi ini, penulis menggunakan metode algoritma Naive Bayes adalah bentuk algoritma berdasarkan metode klasifikasi yang diusulkan oleh ilmuwan Inggris dan disebut teorema Bayes karena memprediksi kemungkinan masa depan berdasarkan pengalaman masa lalu. Teorema ini digabungkan dengan pendekatan naive bayes dimana kondisi dari berbagai properti saling bebas satu sama lain. Pengklasifikasi Naive Bayes mengacu pada keberadaan fitur tertentu dari kelas yang independen dari fitur kelas lainnya (Iwan & Haviani, 2020).

Pengolahan data *mining* menggunakan metode algoritma *Naive Bayes* untuk memprediksi pembelian barang yang laku dan tidak laku. Sehingga penulis tertarik untuk mengangkat judul “**PENERAPAN ALGORITMA NAIVE BAYES DALAM MEMPREDIKSI PENJUALAN MATERIAL BANGUNAN**”.

1.2 Identifikasi Masalah

Berikut ini penulis mencoba untuk uraikan permasalahan pada penelitian ini sebagai berikut:

1. Omset penjualan semakin menurun dengan adanya *kompetitor* yang bergerak dibidang yang sama khususnya dalam bidang material.
2. Data penjualan hanya disimpan tanpa evaluasi lebih lanjut. sehingga sulit dalam memprediksi penjualan barang terlaku dan tidak laku.

3. Permasalahan informasi penjualan bangunan di PT.Tanjung Uncang tidak memakai data *mining* sehingga tidak mudah memprediksikan hasil penjualan kedepannya.

1.3 Batasan Masalah

Agar pembahasan masalah yang ada tetap terfokus pada identifikasi masalah di atas, maka penulis membatasi masalah yang ada dapat diuraikan sebagai berikut:

1. Penelitian ini dilakukan di PT.Tanjung Uncang.
2. Penelitian ini hanya memprediksi penjualan barang material di PT.Tanjung Uncang dengan menggunakan aplikasi bantuan yaitu aplikasi *WEKA*.
3. Data yang diambil oleh peneliti yaitu data 3 tahun dari tahun 2019 sampai dengan tahun 2021, pengambilan data dilakukan langsung dari PT Tanjung Uncang.

1.4 Rumusan Masalah

Berikut ini penulis mencoba merumuskan masalah penelitian ini sebagai berikut:

1. Bagaimana penerapan data *mining* dalam memprediksi penjualan barang material pada PT.Tanjung Uncang menggunakan *Naïve Bayes*?
2. Bagaimana cara implementasikan aplikasi *WEKA* 3.9 dalam memprediksikan penjualan barang material PT.Tanjung Uncang?

1.5 Tujuan Penelitian

Berikut ini penulis dapat memenuhi tujuan penelitian dari skripsi ini sebagai berikut:

1. Untuk mengetahui hasil penerapan data *mining* di PT.Tanjung Uncang dalam memprediksi penjualan bahan material yang laku dan tidak laku.
2. Untuk mengetahui cara implemmentasi data *mining* ke aplikasi *WEKA* dalam memprediksi penjualan yang laku dan tidak laku.

1.6 Manfaat Penelitian

Berikut ini pembuatan skripsi ini diharapkan dapat memberi manfaat sebagai berikut:

1.6.1 Manfaat Teoritis

Secara teoritis, penelitian ini berguna untuk meningkatkan pengalaman, pemahaman, serta wawasan pengetahuan yang lebih luas mengenai algoritma *Naïve Bayes*, sehingga penelitian ini dapat menjadi referensi bagi peneliti lainnya.

1.6.2 Manfaat Praktis

1. Dalam menerapkan ilmu mengenai *Naïve Bayes* dapat menambah pengetahuan dalam memprediksi barang laku dan tidak laku bagi siapa yang akan mengambil pembelajaran tentang data *mining*.

2. Hasil penelitian yang dilakukan dapat diterapkan oleh pengusaha yang ingin mengelolah data agar bisa dimanfaatkan secara *maximal* untuk meningkatkan penjualan.
3. Diharapkan penelitian ini dapat diterapkan pada penelitian lain yang berkaitan dengan *prediksi* barang yang laku dan tidak laku untuk membantu kelancaran kegiatan usaha.

BAB II

KAJIAN PUSTAKA

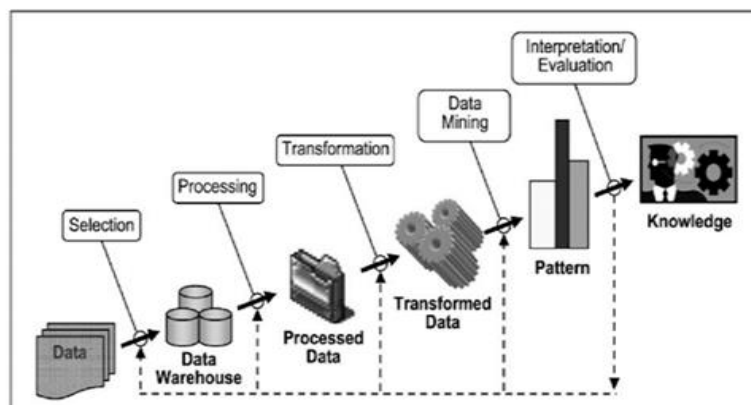
2.1 Konsep Teoritis

Penelitian ini berlandaskan teori-teori yang akan dimulai dari *Knowledge Discovery in Database* (KDD).

2.1.1 *Knowledge Discovery in Database* (KDD)

Knowledge Discovery in Database (KDD) adalah proses untuk menentukan informasi yang bermanfaat seperti berbentuk suatu data yang bersifat baru dan bisa digunakan (Gukguk & Sitohang, 2021). Data informasi yang berukuran besar berguna untuk pola-pola yang ada dalam data dan potensinya yang bermanfaat. *Data mining* merupakan salah satu langkah dari serangkaian proses iterative KDD.

Menurut (Ikhwan et al., 2015) langkah-langkah proses *Knowledge Discovery in Database* (KDD) yang dilakukan diawali pada gambar 2.2 sebagai berikut:



Gambar 2.1 *Knowledge Discovery in Database* (KDD)

1. *Data selection* (seleksi data)

Sebelum tahap *information mining* KDD dimulai, data perlu diseleksi dari sekumpulan data operasional. Data yang dipilih akan diproses dan data disimpan dalam file.

2. *Pre-processing* / pembersihan

Sebelum menjalankan proses data *mining*, harus melalui proses pembersihan data yang menjadi perhatian KDD. Proses pembersihan secara khusus meliputi deduplikasi, pengecekan data yang tidak konsisten, dan perbaikan kesalahan data.

3. Transformasi

Proses transformasi untuk data yang dipilih. Hal ini membuat data cocok untuk proses data *mining*. Proses pengkodean KDD adalah proses kreatif yang bergantung pada jenis dan pola informasi yang di cari di database.

4. *Data mining*

Proses menemukan pola atau informasi yang menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Pilihan metode dan algoritme yang benar sangat berpengaruh pada tujuan proses KDD secara semuanya.

5. interpretasi /evaluasi

Pola informasi yang dihasilkan oleh proses data *mining* harus ditampilkan dalam format yang mudah dipahami oleh para pemangku kepentingan. Tahap ini merupakan bagian dari proses KDD dan disebut interpretasi.

2.2 Data Mining

2.2.1 Pengenalan Data Mining

Data *mining* atau penambangan data adalah proses mencari informasi penting dari data besar. Proses penambangan data selalu menggunakan teknik statistik dan matematika untuk memanfaatkan manfaat teknik kecerdasan buatan (AI) dan penggunaan data dengan hubungan yang tidak terduga (Sikumbang, 2018). Secara sederhana dapat dipahami bahwa data *mining* adalah rangkaian proses untuk mengekstraksi pola-pola menarik dari sekumpulan besar data berupa pengetahuan yang tidak diketahui buatan dan munculnya data mining karena hingga bertahun-tahun pengumpulan akumulasi data seperti data penjualan, data pembelian, data pelanggan, data transaksi, dll (Sinaga & Handoko, 2021). Tujuan dari data *mining* untuk dapat mempelajari lebih lanjut tentang perilaku data yang diamati, atau yang biasa disebut dengan deskripsi tentang apa yang terjadi, untuk dapat memprediksi situasi masa depan.

2.2.2 Proses Data Mining

Proses data *mining* menurut (Budiantara & Budihartanti, 2020) yang dilakukan diawali dari:

1. Description

Identifikasi pola berulang dalam data dan pola tersebut ke dalam aturan dengan standar yang mudah dipahami. Aturan yang dihasilkan harus mudah dipahami agar dapat meningkatkan tingkat pengetahuan secara efektif dalam sistem.

2. *Prediction*

Memprediksi penurunan pelanggan jangka pendek dan mengkategorikannya berdasarkan perilaku masa depan yang diharapkan.

3. *Estimate*

Membangun model dengan catatan lengkap yang memberikan nilai variabel target sebagai nilai prediksi. Selain itu, pada ulasan selanjutnya, estimasi nilai variabel target didasarkan pada nilai variabel prediktor. Misalnya, kami akan memperkirakan tekanan darah sistolik pada pasien rawat inap berdasarkan usia pasien, jenis kelamin, berat badan, dan kadar natrium darah. Hubungan tekanan darah sistolik dengan nilai prediktor selama pembelajaran menghasilkan model perkiraan.

4. *Classification*

Proses menemukan model fungsional dan menggambarkan data sebagai kelas melibatkan pemeriksaan karakteristik objek dan menetapkan objek ke salah satu kelas yang telah ditentukan.

5. *Clustering*

Mengelompokkan data daripada berdasarkan kelas objek tertentu. Tujuannya yaitu menghasilkan pengelompokan objek dalam kelompok yang mirip satu sama lain. Semakin besar kemiripan objek dalam suatu *cluster*, maka semakin besar pula perbedaan antar setiap *cluster*, dan semakin baik kualitas analisis *cluster*.

6. Association

Menemukan satu kemunculan atribut dalam dunia bisnis sering disebut sebagai market basket analysis. Tugas asosiasi mencoba mengungkap aturan yang mengukur hubungan antara dua atau lebih atribut.

2.3 Metode Data Mining

Algoritma *Naive Bayes* adalah cara langsung memprediksikan pola data menggunakan ramalan penjualan bulan depannya. Oleh karena itu, *Naive Bayes* adalah algoritma yang baik untuk meramalkan penjualan (Wijaya & Dwiasnati, 2020). Peneliti menggunakan metode algoritma *Naive Bayes* merupakan metode statistik yang berdasarkan pengenalan pola, menggunakan *Probability* dan biaya yang dihasilkan dari keputusan ini untuk menyelidiki proses klasifikasi disebut *Teorema Bayes* (Fithri, 2017).

Pengklasifikasi *Probability* dengan mudah menghitung satu set probabilitas dan kombinasi beberapa level atau hasil dari kumpulan data yang diperoleh. Kedekatan dan kecepatan *Nave Bayesian* dalam database data besar (Saleh, 2015). Berikut rumus 2.1 *Probability Bayes*:

$$p(x|y) = \frac{p(x \cap y)}{p(y)} \quad \textbf{Rumus 2.1 Probability Bayes}$$

Probability X pada *Y* artinya *Probability* interaksi *X* dan *Y* dari probabilitas *Y*. Atau disebut dengan $P(X|Y)$ artinya persentase banya nya *X* didalam *Y*. Berikut rumus 2.2 *Teorema Bayes*:

$$P(H|X) = \frac{p(X|H)p(H)}{p(X)} \quad \textbf{Rumus 2.2 Teorema Bayes}$$

Penjelasan rumus 2.2 X artinya ciri-ciri, H artinya hipotesis, $P(H|X)$ artinya *Probability* bahwa hipotesis H benar untuk ciri X atau disebut dengan $P(H|X)$ adalah *Probability posterior* H dengan ketentuan X, $P(X|H)$ artinya *Probability* bahwa ciri X benar untuk hipotesis H atau *Probability posterior* X dengan ketentuan H, $P(H)$ adalah *Probability prior* hipotesis H, dan $P(X)$ artinya *Probability prior* bukti X. Berikut Rumus 2.3 *Probability Bayes*:

$$P(Y) = \frac{P(X \cap Y)}{P(Y)} \quad \textbf{Rumus 2.3 Probability Bayes}$$

Probability X didalam Y adalah *Probability* interseksi X dan Y dari *Probability* Y, atau disebut dengan $P(X|Y)$ artinya tingkatan banyaknya X didalam Y. Berikut rumus 2.4 *Teorema bayes*:

$$P(X) = \frac{p(H) p(H)}{p(X)} \quad \textbf{Rumus 2.4 Teorema Bayes}$$

Penjelasan diatas X artinya ciri-ciri, H artinya hipotesis, $P(H|X)$ artinya *Probability* bahwa hipotesis H benar untuk ciri x atau disebut dengan $P(H|X)$ artinya *Probability posterior* H dengan ketentuan X, $P(X|C)$ artinya *Probability*.

2.4 Software Pendukung

Dalam penelitian data mining dapat menggunakan berbagai macam software agar membantu menyokong data mining dalam proses prediksi penjualan bahan material, maka peneliti akan menggunakan software pendukung yaitu *Waikato Knowledge Analysis Environment (WEKA)* merupakan aplikasi penelitian milik

New Zealand yang telah dioleahkan di *University Waikato Trust*. *WEKA* dapat memecahkan masalah data mining dunia nyata, terutama masalah klasifikasi dengan metode pembelajaran mesin. *WEKA* adalah software untuk memahami konsep data mining, yang menyediakan berbagai metode data mining, dimulai dengan mengeksplorasi analisis data, pembelajaran statistik, pembelajaran mesin, dan database. Tidak seperti kebanyakan perangkat lunak penambangan data, *WEKA* didasarkan pada perangkat lunak sumber terbuka, setiap orang dapat mengakses kode sumber dan menambahkan algoritmenya sendiri selama dia setuju dan menyetujui lisensi distribusi perangkat lunak (Purnamasari et al., 2013).

2.5 Penelitian Terdahulu

Penelitian ini tidak akan terwujud tanpa adanya jurnal ataupun sumber pendukung, dan beberapa topik yang berkaitan dengan judul penelitian yang di gunakan sebagai bahan referensi dalam penulisan skripsi ini adalah sebagai berikut:

1. (Nurajijah, 2019) **Algoritma Naïve Bayes, Decision Tree, dan SVM untuk Klasifikasi Persetujuan Pembiayaan Nasabah Koperasi Syariah**. e-ISSN: 2338-0403 dengan menyatakan bahwa keputusan untuk mengotorisasi pembiayaan koperasi Syariah datang dengan risiko gagal bayar yang tinggi ketika utang kredit pelanggan jatuh tempo atau dikenal sebagai kredit macet. Mempertahankan dan meminimalkan risiko ini memerlukan metode yang akurat untuk menentukan persetujuan pendanaan. Tujuan dari penelitian ini

adalah untuk mengklasifikasikan data histori pinjaman nasabah koperasi syariah menggunakan algoritma Naive Bayes, pohon keputusan dan SVM untuk memprediksi keandalan calon nasabah selanjutnya. Hasil penelitian menunjukkan keakuratan algoritma Naive Bayes.

2. (Wijaya & Dwiasnati, 2020) **Implementasi Data Mining dengan Algoritma Naive Bayes pada Penjualan Obat.** ISSN: 2355-6579 menyatakan bahwa tujuan penelitian ini untuk Menganalisis masalah penentuan produk vitamin mana yang dapat atau tidak dapat dijual berdasarkan kategori dapat menjadi panduan bagi apotek untuk menentukan berapa banyak persediaan yang harus mereka miliki di gudang farmasi. Informasi yang diharapkan dari penelitian ini adalah dengan menggunakan algoritma klasifikasi data mining yaitu algoritma Naive Bayes, untuk mendapatkan nilai akurasi data penjualan obat khususnya untuk jenis vitamin yang sering dipilih oleh pelanggan yang membutuhkan obat.
3. (Mustafa et al., 2017) **Implementasi Data Mining untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier.** ISSN: 2460-4259 menyatakan bahwa penelitian ini mengevaluasi kinerja akademik mahasiswa STMIK Dipanegara Makassar selama dua tahun pertama menggunakan teknik data mining algoritma *Naive Bayes Classifier* (NBC), sehingga diklasifikasikan kelulusannya dan memberikan rekomendasi kelulusan mata kuliah tepat waktu. Berdasarkan sejarah kelas yang telah diambil siswa, nilai optimal adalah yang paling tepat.

4. (Annur, 2018) **KLASIFIKASI MASYARAKAT MISKIN MENGGUNAKAN METODE NAÏVE BAYES**. e-ISSN: 2548-7779 menyatakan bahwa masalah utama dalam pengentasan kemiskinan saat ini terkait dengan distribusi pertumbuhan ekonomi yang tidak merata. Dalam penelitian ini, kami menggunakan teknik data mining untuk mengklasifikasikan data penduduk miskin yang diperoleh di wilayah Tibawa. Metode yang digunakan adalah metode classifier Naive Bayes yang merupakan salah satu teknik klasifikasi data mining. Hasil pengujian tabel klasifikasi sore menggunakan metode split-validation akurat 73% pada dataset yang diteliti menggunakan metode klasifikasi Naive Bayes, atau termasuk dalam kategori Baik. Nilai presisi adalah 92% dan tingkat recall adalah 86%.
5. (Nofriansyah et al., 2016) **Penerapan Data Mining dengan Algoritma Naive Bayes Clasifier untuk Mengetahui Minat Beli Pelanggan terhadap Kartu Internet XL (Studi Kasus di CV. Sumber Utama Telekomunikasi)**. ISSN: 1978-6603 menyatakan bahwa persaingan terjadi dalam dunia bisnis, dan pelaku harus terus mempertimbangkan strategi dan kemajuan untuk menjamin kelangsungan bisnis yang mereka operasikan. Hal ini menyebabkan persaingan di antara penyedia kartu untuk kartu internet. Penyedia kartu internet berlomba-lomba menarik pelanggan melalui berbagai strategi pemasaran. Untuk bertahan hidup tanpa mengurangi daya saing, metode klasifikasi dapat menemukan model yang membedakan konsep atau kelas data. Tujuannya adalah untuk dapat

menyimpulkan kelas objek dengan label yang tidak diketahui sehingga algoritma Naive Bayes dapat memprediksi prospek masa depan berdasarkan pengalaman sebelumnya. Hasil penelitian ini membantu perusahaan memprediksi atau memprediksi perilaku kartu internet baru. Kemampuan untuk membuat keputusan dan meningkatkan strategi pemasaran.

6. (Putro et al., 2020) **Penerapan Metode Naive Bayes Untuk Klasifikasi Pelanggan.** e-ISSN: 2620-7532 menyatakan bahwa lokasi usaha memegang peranan penting dalam penjualan. Perusahaan ini berbasis di kota, sehingga memudahkan vendor untuk mendistribusikan acara kepada orang-orang. Kegiatan distribusi erat kaitannya dengan kegiatan penjualan. Jika ada penawaran penjualan untuk pelanggan potensial dan kelompok pelanggan non-pilihan. Salah satu metode yang dapat digunakan untuk klasifikasi adalah data mining. Penambangan data kategorikal yang paling umum digunakan adalah metode Naive Bayes. Atribut yang digunakan dalam proses klasifikasi pelanggan adalah jumlah pembelian, jangka waktu, dan lokasi. Sistem klasifikasi menghasilkan 23 jawaban valid dan 2 jawaban salah. Menurut hasil yang diperoleh dengan metode confusion matrix, tingkat akurasi adalah 92%, tingkat akurasi 100%, dan tingkat recall 91%.
7. (Kurniawan, 2018) **PERBANDINGAN ALGORITMA NAIVE BAYES DAN C.45 DALAM KLASIFIKASI DATA MINING.** e-ISSN: 2528-6579 menyatakan bahwa metode Naive Bayes dan C.45 diterapkan pada 4 studi kasus, yaitu kasus penerimaan “Kartu Indonesia Sehat”, identifikasi pengajuan kartu kredit di bank, penentuan usia saat lahir, dan penentuan

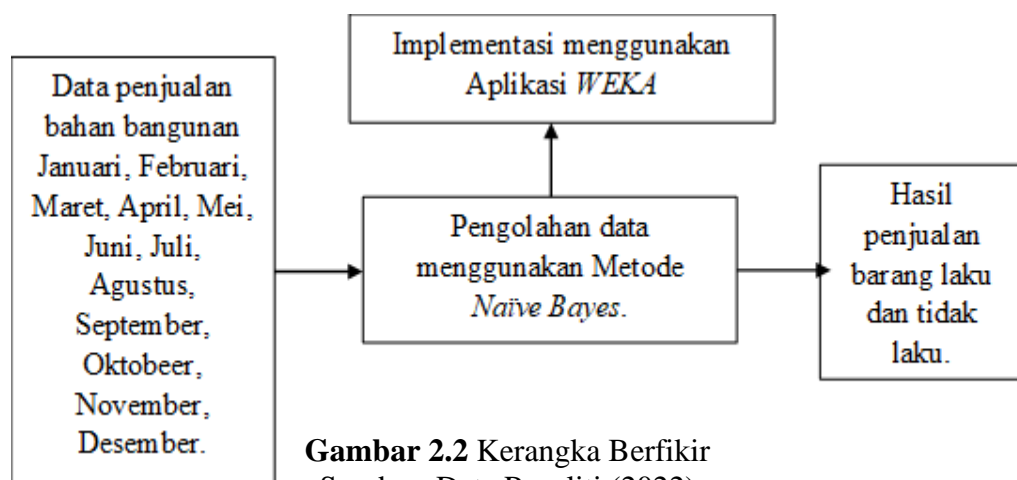
kelayakan. Temukan algoritma terbaik dalam koperasi, dalam setiap kasus. Precision, recall, dan precision kemudian dibandingkan untuk setiap data training dan test. Berdasarkan hasil implementasi, telah dibangun sebuah aplikasi untuk menerapkan algoritma Naive Bayes dan C.45 pada 4 kasus tersebut. Aplikasi diuji dengan kotak hitam dan algoritme dan hasilnya berhasil dan dapat menjalankan kedua algoritme dengan benar. Berdasarkan hasil pengujian, semakin banyak data pelatihan yang digunakan, semakin tinggi nilai presisi dan recall.

8. (K.Vembandasamy et al., 2015) **Heart Diseases Detection Using Naive Bayes Algorithm**. ISSN 2348 – 7968 menyatakan bisnis kesehatan telah menjadi bidang penting dari berbagai obat-obatan dalam industri kesehatan mengandung banyak data dan informasi tersembunyi untuk itu informasi tersembunyi ini akan digunakan untuk membuat keputusan yang efektif dengan menerapkan teknik data mining. Data mining dapat memecahkan banyak masalah kesehatan. Algoritma Naive Bayes mewakili metode penambangan data untuk mendiagnosis penyakit jantung, dan mengusulkan sistem prediksi penyakit jantung (HDPS) berdasarkan metode penambangan data dengan menganalisis beberapa parameter untuk memprediksi penyakit jantung.
9. (Peling et al., 2017) **Implementation of Data Mining To Predict Period of Students Study Using Naive Bayes Algorithm**. e-ISSN: 2579-597X menyatakan kualitas perguruan tinggi di Indonesia, khususnya kualitas program sarjana, diukur dari akreditasi yang dilakukan mahasiswa yang

tidak lulus tepat waktu masih menjadi perbincangan hangat tentang kegagalan akademik. Ketepatan waktu mahasiswa sarjana dapat digunakan untuk menentukan pola kelulusan mahasiswa sarjana melalui teknik data mining yang dapat digunakan sebagai dasar untuk memprediksi kelulusan mahasiswa tahun depan. Studi ini menunjukkan bahwa *Naive Bayes* dapat mengklasifikasikan data uji dengan benar dengan tingkat kesalahan rata-rata masing-masing 86,16% dan 13,84%.

2.6 Kerangka Berfikir

Setelah memaparkan teori-teori di atas, penelitian ini menggunakan kerangka pemikiran yang dirancang oleh peneliti seperti gambar 2.2 berikut:



Gambar 2.2 Kerangka Berfikir
Sumber: Data Peneliti (2022)

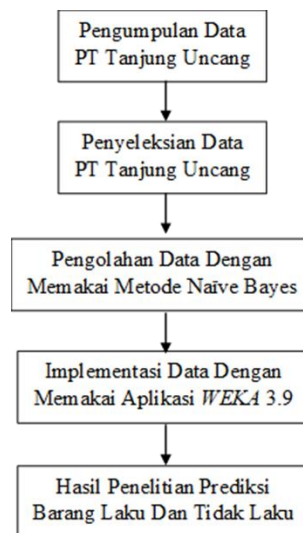
Data penjualan material, diambil dari hasil data arsip PT Tanjung Uncang yang digunakan sebagai masukan dalam penelitian, lalu proses pengolahan datanya memakai metode *Naive Bayes* dengan mengimplementasikan ke aplikasi *WEKA*. Keluaran hasilnya berupa barang laku dan tidak laku.

BAB III

METODE PENELITIAN

3.1 Desain Penelitian

Dalam perancangan desain penelitian memerlukan sebuah desain sebagai arah penjas, dimulai dengan cara memperoleh data, mengolah data, dan langkah-langkah untuk menyelesaikan pengolahan data. Berikut ini adalah penjelasan dan desain penelitian yang dirancang oleh peneliti, berikut gambar 3.1:



Gambar 3.1 Desain Penelitian
Sumber: Data Penelitian (2022)

Berdasarkan desain penelitian pada gambar 3.1, dapat dijelaskan sebagai berikut:

1. Pengumpulan Data

Pengumpulan data adalah untuk memperoleh informasi yang diperlukan untuk mencapai tujuan penelitian. Pengumpulan data dilakukan di PT Tanjung Uncang dengan menggunakan teknik wawancara, observasi dan

mencari referensi terpercaya yang berkaitan dengan penelitian data mining dan materi *Naive Bayes* yang diperoleh seperti *e-book* dan jurnal.

2. Penyeleksian Data

Tujuan pemilihan data peneliti adalah untuk memperoleh data yang benar-benar memenuhi tujuan penelitian untuk dijadikan sampel. Data yang dipilih berupa data penjualan barang material pada PT Tanjung Uncang 2019 sampai dengan tahun 2021.

3. Pengolahan Data

Metode *Naive Bayes* mengolah data yang sudah dipilih dan kemudian mengolahnya menjadi teknik data *mining*.

4. Implementasi

Implementasi *Naive Bayes* digunakan ke aplikasi *WEKA* 3.9 pada penelitian ini untuk memprediksi tingkat penjualan PT Tanjung Uncang.

5. Hasil Penelitian

Hasil menunjukkan penelitian yang telah dilakukan dengan menggunakan metode *Naive Bayes* teknik komputasi data *mining* dan diuji dengan aplikasi *WEKA* 3.9, hasilnya memprediksi penjualan barang material yang laku dan tidak laku PT Tanjung Uncang.

3.2 Teknik Pengumpulan Data

Adapun teknik pengumpulan data yang digunakan peneliti dalam penelitian ini adalah sebagai berikut:

1. Observasi

Teknik pengumpulan data observasi adalah rangkaian proses mental dan fisik yang kompleks. Observasi dalam penelitian ini adalah untuk mendapatkan data yang dibutuhkan dengan cara observasi langsung ke PT Tanjung Uncang yaitu data penjualan barang material supaya mengetahui variabel yang dibutuhkan dalam proses pengolahan data.

2. Wawancara

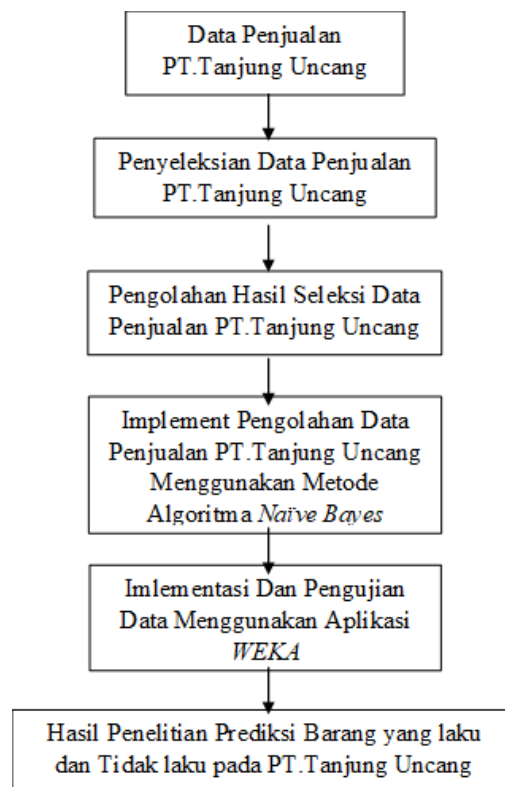
Teknik pengumpulan data dengan wawancara adalah proses mendapatkan informasi langsung seperti teknik interview. Dalam penelitian ini peneliti melakukan tanya jawab dengan pihak PT Tanjung Uncang untuk mendapatkan informasi mengenai penjualan barang material dan apa saja kendala yang menyangkut masalah penelitian saat melakukan pembelian bahan bangunan di PT Tanjung Uncang.

3.3 Operasional Variabel

Proses variabel penelitian memanfaatkan hasil jualan bahan material pada PT Tanjung Uncang. Untuk mendukung penelitian ini, peneliti mengambil hasil penjualan data 3 tahun dari tahun 2019 sampai tahun 2021 dan variabelnya yaitu bulan Januari, February, Maret, April, Mei, Juni, Juli, Agustus, September, Oktober, November, dan Desember setiap tahun selama 3 tahun terakhir, yaitu 2019 hingga 2021.

3.4 Metode Perancangan Sistem

Pada pembuatan skripsi ini dibutuhkan perancangan pada penelitian dapat dilihat gambar 3.2 sebagai berikut:



Gambar 3.2 Metode Perancangan Sistem
Sumber: Data Peneliti (2022)

Berdasarkan metode perancangan sistem pada gambar 3.2 dapat dijelaskan sebagai berikut:

1. Data Penjualan

Data penjualan adalah data yang dikumpulkan berupa data transaksi antara toko bangunan memesan barang material ke distributor PT.Tanjung Uncang dan data yang diperoleh selama 3 tahun akan diterapkan ke data mining.

2. Penyeleksian Data

Melakukan Peneyelesian data penjualan yang telah diperoleh perlu dilakukan seleksi sampai menghasilkan data yang layak dan membersihkan data yang tidak konsisten.

3. Hasil Seleksi Data

Data penjualan yang selesai diseleksi akan diolahkan ke bentuk Teknik data mining.

4. Implementasi Data Penjualan Ke Algoritma *Naïve Bayes*

Pengolahan data penjualan yang sudah selesai diseleksikan akan pilih dan kemudian akan di implemementasi ke algoritma *Naïve Bayes*.

5. Implementasi dan Pengujian Aplikasi *WEKA*

Mengimplementasikan algoritma *Naïve Bayes* ke dalam pengujian Aplikasi *WEKA* untuk memprediksi penjualan material.

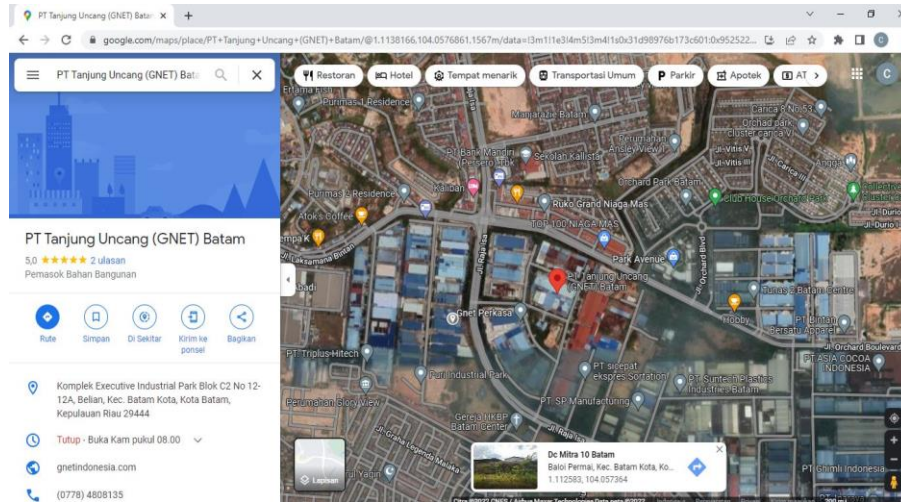
6. Hasil Penelitian

Aplikasi yang dijalankan oleh *WEKA* digunakan sebagai pembuktian apakah hasil yang diperoleh dari proses perhitungan.

3.5 Lokasi dan Jadwal Penelitian

Penelitian ini dilakukan oleh penulis pada PT. Tanjung Uncang yang berlokasi di daerah Batam Center, Kawasan Executive Industrial Park Blok C2 No

12-12A. Untuk waktu penelitian ini dimulai pada bulan Maret 2022 sampai bulan Juli 2022.



Gambar 3.3 Lokasi PT.Tanjung Uncang
Sumber: Data Peneliti(2022)

Tabel 3.1 Jadwal Penelitian

Kegiatan	Mar 2022				Apr 2022				Mei 2022				Jun 2022				Jul 2022			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Persiapan Penelitian	■	■																		
Identifikasi data			■	■																
Pengumpulan Data					■	■	■	■												
Hitungan manual									■	■	■	■								
Aplikasi WEKA 3.9													■	■	■	■				
Menulis dan Upload																	■	■	■	■

Sumber: Data Peneliti (2022)