

**IMPLEMENTASI *DATA MINING K-NEAREST  
NEIGHBOUR* PADA PENERIMAAN  
KARYAWAN DI  
PT. DWI SUMBER ARCA WAJA**

**SKRIPSI**



**Oleh:  
Ega Saputra  
180210122**

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

**IMPLEMENTASI *DATA MINING K-NEAREST  
NEIGHBOUR* PADA PENERIMAAN  
KARYAWAN DI  
PT. DWI SUMBER ARCA WAJA**

**SKRIPSI**

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



**Oleh  
Ega Saputra  
180210122**

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

## SURAT PERNYATAAN ORISINALITAS

Yang bertanda tangan di bawah ini saya:

Nama : Ega Saputra  
NPM : 180210122  
Fakultas : Teknik dan Komputer  
Program Studi : Teknik Informatika

Menyatakan bahwa “Skripsi” yang saya buat dengan judul:

**IMPLEMENTASI DATA *MINING K-NEAREST NEIGHBOUR* PADA  
PENERIMAAN KARYAWAN DI PT DWI SUMBER ARCA WAJA**

Adalah hasil karya sendiri dan bukan “duplikasi” dari karya orang lain. Sepengetahuan 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 diproses sesuai dengan peraturan perundang-undangan yang berlaku.

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

Batam, 21 Januari 2022



**Ega Saputra**  
NPM. 180210122

**IMPLEMENTASI *DATA MINING K-NEAREST  
NEIGHBOUR* PADA PENERIMAAN  
KARYAWAN DI  
PT. DWI SUMBER ARCA WAJA**

**SKRIPSI**

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

**Oleh:  
Ega Saputra  
180210122**

**Telah disetujui oleh Pembimbing pada tanggal  
seperti tertera dibawah ini**

**Batam, 24 Januari 2022**



**Rahmat Fauzi, S.Kom., M.Kom.  
Pembimbing**



---

**Universitas Putera Batam**

## ABSTRAK

Perekrutan karyawan merupakan langkah awal yang dilakukan oleh PT. Dwi Sumber Arca Waja (PT. DSAW) untuk merekrut calon tenaga kerja profesional. Profesionalisme diperlukan sebagai salah satu input untuk menghasilkan output yang sesuai. PT. DSAW adalah perusahaan manufaktur Pipa Baja Besar yang bertindak sebagai pemasok yang menyediakan barang untuk mendukung sektor minyak dan gas. Kantor Pusat berlokasi di Jl. Hang Kesturi I No.8, Kawasan Industri, Batam. Seiring berjalannya waktu, PT. DSAW perlu mencari cara agar lebih cepat dan efisien dalam menentukan calon karyawan sehingga dapat dengan cepat menemukan karyawan yang cocok dan sesuai. Sehingga pada penelitian ini ditujukan untuk menerapkan metode klasifikasi *K-Nearest Neighbour* (K-NN) pada sistem seleksi rekrutmen di PT. DSAW. Metode K-NN akan menghitung tingkat kemiripan dengan mengukur jarak antara kebutuhan yang ditetapkan oleh PT. DSAW dengan data yang dimiliki oleh calon karyawan. *Data set* yang digunakan untuk mencari nilai prediksi dengan keakuratan yang tinggi di ambil dari data penerimaan karyawan dari tahun 2018 sampai dengan 2020. Kemudian didapatkan tingkat keakuratan terhadap klasifikasi *K-Nearest Neighbour* yaitu 100%. Sehingga dapat di ambil kesimpulan yaitu cara menganalisis data karyawan dengan menggunakan algoritma *K-Nearest Neighbour* ini yaitu dengan melakukan seleksi pada data uji, *preprocessing*, *transformation*, menentukan jarak, kemudian menggunakan rangking. Kemudian data hasil *transformation* karyawan di PT. DSAW diolah kedalam aplikasi *RapidMiner* berbasis algoritma *K-Nearest Neighbour* dan menghasilkan nilai akurasi terhadap klasifikasi algoritma *K-Nearest Neighbour* yaitu 100%. Dimana prediksi departemen terbanyak terdapat pada departemen *Accounting*, *Com & Log*, *Engineer*, *HRD & GA*, *HSE*, *PPC*, *QA*. Dalam menerapkan klasifikasi algoritma *K-Nearest Neighbour* dalam penentuan calon karyawan, dilakukan dengan cara mengelola data yang ada dan melakukan perhitungan manual dan uji menggunakan aplikasi bantu *data mining* yaitu *RapidMiner*.

Kata Kunci : *Data Mining*, *K-Nearest Neighbour*, *Recruitment*, PT. Dwi Sumber Arca Waja

## **ABSTRACT**

*Employee recruitment is the first step taken by PT. Dwi Sumber Arca Waja (PT DSAW) to recruit prospective professional workers. Professionalism is needed as one of the inputs to produce the appropriate output. PT. DSAW is a Large Steel Pipe manufacturing company that acts as a supplier providing goods to support the oil and gas sector. The Head Office is located on Jl. Hang Kesturi I No.8, Industrial Estate, Batam. Over time, PT. DSAW needs to find a way to more quickly and efficiently determine prospective employees so that they can quickly find suitable and suitable employees. So that this study is aimed at applying the K-Nearest Neighbour (K-NN) classification method to the recruitment selection system at PT. DSAW. The K-NN method will calculate the level of similarity by measuring the distance between the needs set by PT. DSAW with data held by prospective employees. The data set used to find the predictive value with high accuracy is taken from employee recruitment data from 2018 to 2020. Then the accuracy level of the K-Nearest Neighbour classification is 100%. So it can be concluded that how to analyze employee data using the K-Nearest Neighbour algorithm, namely by selecting test data, preprocessing, transformation, determining distance, then using ranking. Then the data from the transformation of employees at PT. DSAW is processed into the RapidMiner application based on the K-Nearest Neighbour algorithm and produces an accuracy value of the K-Nearest Neighbour algorithm classification, which is 100%. Where the most departmental predictions are in the departments of Accounting, Com & Log, Engineer, HRD & GA, HSE, PPC, QA. In applying the classification of the K-Nearest Neighbour algorithm in determining prospective employees, it is done by managing existing data and performing manual calculations and testing using a data mining aid application, namely RapidMiner.*

*Keywords: Data Mining, K-Nearest Neighbour, Recruitment, PT. Dwi Sumber Arca Waja*

## KATA PENGANTAR

Segala Puji dan Syukur penulis panjatkan atas kehadiran Allah SWT yang dengan senantiasa telah mencurahkan berkat dan karunia-Nya, sehingga penulis bisa menyusun dan menyelesaikan tugas akhir yang berjudul “Implementasi *Data Mining K-Nearest Neighbour* Pada Penerimaan Karyawan di PT. Dwi Sumber Arca Waja”

Penulisan tugas akhir ini disusun guna memenuhi persyaratan dalam menyelesaikan program studi Strata Satu (S1) pada program Studi Teknik Informatika Universitas Putera Batam.

Penulis sadar bahwa tugas akhir ini tidak dapat terselesaikan dengan baik tanpa dukungan, bimbingan serta bantuan dari berbagai pihak. Untuk itu pada kesempatan ini dengan segala kerendahan hati, penulis mengucapkan terima kasih yang sebesar-besarnya kepada yang terhormat:

1. Ibu Dr. Nur Elfi Husda, S.Kom., M.SI. Selaku Rektor Universitas Putera Batam.
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. Kepada Bapak Rahmat Fauzi, S.Kom., M.Kom. Selaku pembimbing Skripsi pada Program Studi Teknik Informatika Universitas Putera Batam.
5. Kepada Bapak Cosmas Eko Suharyanto, S.Kom., M.MSI. Selaku pembimbing Akademik pada Program Studi Teknik Informatika Universitas Putera Batam.



6. Para Dosen serta Staff di Universitas Putera Batam.
7. Kedua orang tua, kakak dan saudara yang selalu memberikan dukungan baik dari segi material maupun moril kepada penulis.
8. Seluruh keluarga besar penulis yang selalu memberi nasehat dan motivasi.
9. PT. Dwi Sumber Arca Waja yang telah memberikan izin kepada penulis untuk pengumpulan data penelitian.
10. Para teman-teman seperjuangan yang dengan setia mendampingi dan memberikan dukungan dan juga semangat.

Penulis juga menyadari keterbatasan pengalaman dan juga pengetahuan yang dimiliki penulis, sehingga dalam penulisan tugas akhir ini masih sangat jauh dari kata sempurna. Oleh sebab itu, penulis sangat mengharapkan saran dan kritik yang konstruktif dari berbagai pihak. Akhir kata semoga tugas akhir ini dapat diterima dan bermanfaat bagi para pembaca

Batam, 18 Januari 2022

Penulis,



Ega Saputra

180210122



---

**Universitas Putera Batam**

## DAFTAR ISI

<b>HALAMAN SAMPUL</b> .....	i
<b>HALAMAN JUDUL</b> .....	ii
<b>SURAT PERNYATAAN</b> .....	iii
<b>HALAMAN PENGESAHAN</b> .....	iv
<b>ABSTRAK</b> .....	vi
<b>ABSTRACT</b> .....	vii
<b>KATA PENGANTAR</b> .....	viii
<b>DAFTAR ISI</b> .....	xi
<b>DAFTAR GAMBAR</b> .....	xiii
<b>DAFTAR TABEL</b> .....	xiv
<b>DAFTAR RUMUS</b> .....	xv
<b>BAB I PENDAHULUAN</b> .....	1
1.1 Latar Belakang Penelitian.....	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
<b>BAB II TINJAUAN PUSTAKA</b> .....	8
2.1 <i>Knowledge Discovery in Database (KDD)</i> .....	8
2.2 <i>Data Mining</i> .....	11
2.2.1 Pengelompokan <i>Data Mining</i> .....	12
2.2.2 Prediksi .....	14
2.2.3 <i>Time Series Analisis</i> .....	15
2.2 <i>Metode Data Mining</i> .....	17
2.3 Algoritma <i>K-Nearest Neighbour (K-NN)</i> .....	17
2.4 <i>Software</i> Pendukung .....	20
2.5 Penelitian Terdahulu .....	21
2.6 Kerangka Pemikiran .....	25
<b>BAB III METODE PENELITIAN</b> .....	27
3.1 Desain Penelitian (Arsitektur Sistem) .....	27

3.2	Teknik Pengumpulan Data .....	29
3.3	Operasional Variabel .....	31
3.4	Metode Analisis dan Rancangan Sistem ( <i>Data Mining</i> ) .....	32
3.5	Lokasi Dan Jadwal Penelitian.....	35
3.5.1	Lokasi Penelitian .....	35
3.5.1	Jadwal Penelitian .....	36
	<b>BAB IV HASIL DAN PEMBAHASAN</b> .....	38
4.1	Analisis Data.....	38
4.1.1	Data Perekrutan Karyawan .....	38
4.2	Hasil.....	41
4.3	Pembahasan .....	42
4.3.1	<i>Data Selection</i> .....	42
4.3.2	<i>Preprocessing</i> .....	43
4.3.3	<i>Transformation</i> .....	45
4.3.4	Implementasi <i>RapidMiner</i> .....	48
4.4.4	Susunan Operator Algoritma <i>K-Nearest Neighbour</i> .....	49
	<b>BAB V PENUTUP</b> .....	53
5.1	Kesimpulan .....	53
5.2	Saran .....	54
	<b>DAFTAR PUSTAKA</b> .....	56
	<b>Lampiran</b> .....	60
	Lampiran 1. Pendukung Penelitian .....	60
	Lampiran 2. Daftar Riwayat Hidup.....	75
	Lampiran 3. Surat Keterangan Penelitian .....	76

## DAFTAR GAMBAR

<b>Gambar 1. 1</b> Tahapan Proses KDD .....	9
<b>Gambar 2. 1</b> Logo <i>RapidMiner</i> .....	20
<b>Gambar 3. 1</b> Lokasi Penelitian .....	36
<b>Gambar 4. 1</b> Grafik Data Perekrutan Karyawan .....	40
<b>Gambar 4. 2</b> Grafik Model Sub Proses <i>Clasification K-Nearest Neighbour</i> .....	48
<b>Gambar 4. 3</b> Hasil Perhitungan Nilai Akurasi .....	48
<b>Gambar 4. 4</b> Model Klasifikasi K-NN .....	49
<b>Gambar 4. 5</b> <i>Importing data Training</i> Pada Operator <i>Read Excel</i> .....	50
<b>Gambar 4. 6</b> Susunan Operator algoritma K-NN .....	50
<b>Gambar 4. 7</b> Hasil Prediksi Rekrutment Karyawan .....	51

## DAFTAR TABEL

<b>Tabel 3. 1</b> Sampel Data Perekrutan Karyawan.....	30
<b>Tabel 3. 2</b> Jadwal Penelitian.....	36
<b>Tabel 4. 1</b> Data Perekrutan Karyawan.....	39
<b>Tabel 4. 2</b> Hasil Perekrutan Karyawan.....	40
<b>Tabel 4. 3</b> Hasil Perekrutan Karyawan.....	42
<b>Tabel 4. 4</b> <i>Data Selection</i> Karyawan.....	43
<b>Tabel 4. 5</b> Sampel <i>Data Selection</i> Karyawan Tahun 2020.....	44
<b>Tabel 4. 6</b> Sampel <i>Data Selection</i> Karyawan Tahun 2019.....	44
<b>Tabel 4. 7</b> Sampel <i>Data Selection</i> Karyawan Tahun 2018.....	45
<b>Tabel 4. 8</b> <i>Data Testing</i> Kebutuhan Karyawan Tahun 2018 Sampai 2020.....	46

## DAFTAR RUMUS

<b>Rumus 2. 1</b> Perhitungan Jarak <i>Euclidean</i> .....	19
--	----



**Universitas Putera Batam**



# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang Penelitian**

Sebuah perusahaan tentunya memiliki karyawan atau pegawai, dimana karyawan sendiri adalah komponen penting yang ada pada perusahaan tersebut. Disisi lain, untuk penentuan dan proses seleksi karyawan itu sendiri haruslah memenuhi kriteria sesuai dengan persyaratan yang perusahaan tentukan. Secara umum tahapan pada penerimaan karyawan meliputi *screening* atau seleksi lamaran kerja dan *CV (Curriculum Vitae)*, tes tertulis (logika dan matematika dasar) untuk bagian tertentu, tes *interview*, cek kesehatan, dan terakhir yaitu tanda tangan kontrak. Proses seleksi karyawan di PT. Dwi Sumber Arca Waja (DSAW) pun sedemikian rupa, dimana setiap proses seleksi karyawan tentunya akan membutuhkan waktu agar dapat menemukan calon karyawan yang cocok dan selanjutnya berhak lolos hingga sampai menjadi seorang karyawan.

PT. DSAW adalah perusahaan pemasok migas yang bergerak di bidang *Manufacturing of Large Steel Pipes* dan berlokasi di Jl. Hang Kesturi Km.4, Kawasan Industri Kabil, dengan tujuan memenuhi permintaan sektor minyak dan gas bumi. Perusahaan ini memenuhi kebutuhan *Client* dengan cara memproduksi pipa besar secara terpisah sesuai *project* yang di sepakati dalam sebuah tender kerja. Berangkat dari latar belakang perusahaan dengan sistem proyek, maka setiap ada proyek baru dari *client* pihak HRD (*Human Resources Department*) dari PT. DSAW

akan berusaha untuk memenuhi kebutuhan *manpower* demi menyeimbangkan antara ketersediaan tenaga kerja dan hasil output yang di inginkan.

Dalam hal penyeleksian karyawan baru, ada banyak hal yang harus di pertimbangkan, biasanya pihak HRD pada saat pelaksanaan *screening* lamaran dan CV kerja maka akan di pilih kandidat yang memenuhi kriteria seperti memiliki sertifikat welder untuk bagian welder, memiliki surat pengalaman pekerjaan yang sesuai yang di lamar, kelengkapan berkas dan syarat-syarat pendukung lainnya. Setelah itu, maka akan dilanjutkan ke proses tes *interview* untuk penyeleksian karyawan, di proses ini pihak HRD akan mengalami kesulitan bila menemukan kandidat yang tidak sesuai dengan kecakapan ketika prosesi tes *interview*, tidak memiliki sikap yang baik serta tidak hadirnya calon karyawan ketika proses tes *interview*. Sebagai akibat dari masalah ini, proses seleksi untuk pekerja potensial akan memakan waktu lama, menimbulkan dilema selama perekrutan karyawan. Jika ini tidak segera diperbaiki, itu akan terus terjadi dan dalam waktu jangka panjang data ini akan semakin menumpuk dan hanya akan mejadi arsip perusahaan. Maka dari itu, perlu adanya solusi untuk menghindari terjadinya kesalahan dan juga untuk meminimalisis permasalahan yang ada, peneliti menggunakan metode *data mining*.

*Data mining* adalah proses memilah-milah sejumlah besar data untuk mencari pola dan wawasan yang bermakna. Ini termasuk pengumpulan data, ekstraksi, analisis, dan pengukuran statistik. *Knowledge discovery*, *Knowledge extraction*, *data/pattern analysis*, *information harvesting*, dan istilah lainnya digunakan untuk menggambarkan *data mining* (Arhami and Nasir, 2020). Karena permasalahannya

adalah bagaimana cara untuk memudahkan pihak HRD dalam pengambilan keputusan, maka penulis menggunakan sistem algoritma *K-Nearest Neighbour* (K-NN).

*K-Nearest Neighbour* adalah metode pencarian kasus yang menghitung jarak antara instans baru dan lama berdasarkan kecocokan berat sejumlah atribut yang ada yang sebanding (Sitepu and Buulolo, 2017). *K-Nearest Neighbour* memiliki manfaat milik kelompok klasifikasi data yang sangat dasar, mudah, dan efektif pada data yang sangat besar, dan dapat secara akurat mengkategorikannya. Sehingga dapat digunakan sebagai pengklasifikasi pada penelitian ini.

Penulis berencana untuk melakukan studi dengan judul “**Implementasi *Data mining K-Nearest Neighbour* pada Penerimaan Karyawan di PT. Dwi Sumber Arca Waja**” berdasarkan latar belakang yang diberikan di atas.

## **1.2 Identifikasi Masalah**

Beberapa masalah berikut dapat dideteksi berdasarkan pernyataan yang disebutkan di atas:

1. Proses seleksi calon karyawan yang masih menggunakan system manual sehingga membutuhkan waktu yang tidak efisien
2. Kesulitan dalam penentuan kandidat yang benar-benar sesuai dan kompeten dan layak untuk mejadi karyawan PT. Sumber Arca Waja.

### 1.3 Batasan Masalah

Penulis membatasi masalah yang akan dijelaskan, antara lain, karena masalah yang dinyatakan dalam penelitian ini begitu luas dan agar tidak mengalihkan dari tujuan utama penulis:

1. Implementasi ini dilakukan berdasarkan pada data perekrutan karyawan PT. Dwi Sumber Arca Waja (DSAW).
2. Data yang digunakan dalam masa waktu tiga tahun terakhir yaitu periode Januari 2018 sampai dengan Desember 2020 yang diperoleh dari tim rekrutment PT. DSAW.
3. Data perekrutan karyawan yang digunakan adalah berdasar data perekrutan karyawan perbulan dengan menggunakan metode *K-Nearest Neighbour*.
4. Aplikasi *data mining* yang digunakan untuk dalam pengolahan data adalah *RapidMiner*.

### 1.4 Rumusan Masalah

Dimungkinkan untuk merumuskan topik penelitian yang akan dieksplorasi berdasarkan gambar latar belakang, yaitu:

1. Bagaimana menganalisis data karyawan sehingga menghasilkan data yang akurat ke dalam *RapidMiner* dan menghasilkan prediksi dengan akurasi yang paling tinggi?
2. Bagaimana menerapkan klasifikasi *K-Nearest Neighbour* dalam penentuan atau seleksi calon karyawan PT. DSAW?

## 1.5 Tujuan Penelitian

Tujuan dari penelitian ini, berdasarkan citra formulasi masalah, adalah untuk:

1. Untuk menganalisis data karyawan sehingga menghasilkan data yang akurat ke dalam *RapidMiner* dan menghasilkan prediksi dengan akurasi yang paling tinggi.
2. Untuk mengetahui penerapan klasifikasi *K-Nearest Neighbour* dalam penentuan atau seleksi calon karyawan PT. DSAW.

## 1.6 Manfaat Penelitian

Manfaat dari penelitian dapat diklasifikasikan ke dalam dua kategori: teoritis dan praktis.

### 1.6.1 Manfaat Teoritis

1. Memberikan wawasan terkait system perekrutan beserta persyaratan yang harus di penuhi.
2. Selain berfungsi sebagai referensi bagi pembaca yang ingin mempelajari lebih lanjut tentang algoritma *K-Nearest Neighbour* untuk penambangan data, terutama di bidang Teknik Informatika.

### 1.6.2 Manfaat Praktis

1. Membantu staff dan tim rekrutment dalam perekrutan calon karyawan.
2. Bagi mahasiswa yang ingin melakukan studi lebih lanjut tentang *data mining*, sebagai referensi atau lebih banyak pengetahuan.

- 3 Untuk mendapatkan pemahaman yang lebih baik tentang teknik penambangan data *K-Nearest Neighbour* dan untuk memenuhi salah satu prasyarat untuk gelar sarjana teknik informatika.



---

**Universitas Putera Batam**

## **BAB II**

### **TINJAUAN PUSTAKA**

#### **2.1 *Knowledge Discovery in Database (KDD)***

Pola dan hubungan big data dapat ditemukan melalui penggunaan data historis, teknik yang dikenal sebagai penemuan pengetahuan melalui *data mining* (KDD). KDD sering digunakan untuk menggambarkan proses penyaringan melalui database besar untuk informasi yang relevan. KDD menurut (Mardianti and Fauzi, 2020) merupakan kegiatan yang digunakan dalam pengumpulan data, dan mengolah data untuk menemukan pola hubungan yang teratur dalam basis data yang besar. Saat ini, KDD semakin populer karena penurunan kebutuhan akan pengenalan pola yang telah terjadi sebagai akibat dari perkembangan ini. Hasil penambangan data dapat digunakan di masa depan untuk membuat keputusan yang lebih tepat tentang masa depan. Menurut (Buulolo, 2020) *Knowledge Discovery in Database* (KDD) mencakup proses lengkap mengekstraksi atau menemukan pola, pengetahuan, dan informasi yang berpotensi berguna dari kumpulan data dalam jumlah besar, dimana pengetahuan dan informasi yang dihasilkan dari KDD bersifat sah, baru, mudah di mengerti serta bermanfaat.

Menurut Tomar, Agarwal dalam (Widaningsih, 2019) KDD adalah proses menemukan informasi baru yang berharga yang lebih mudah dipahami daripada sistem penyimpanan data yang besar dan kompleks. Dalam proses KDD, hasil dari kumpulan data ditafsirkan dengan menyatukan kumpulan data dengan informasi

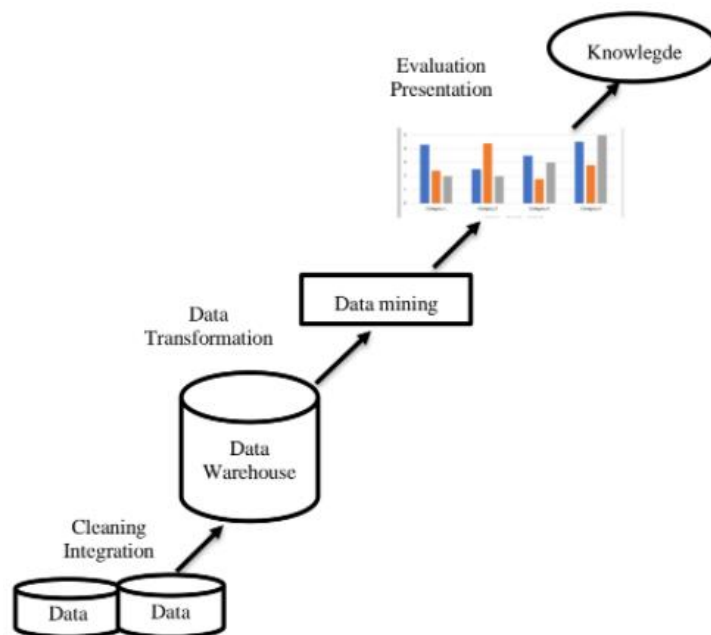


dari bidang lain. Penting untuk dicatat bahwa proses KDD dimulai dan diakhiri dengan penetapan dan evaluasi tujuan.

Atas dasar hal di atas, dapat dinyatakan bahwa KDD adalah teknik untuk menemukan atau mengekstraksi informasi dari toko data besar melalui pengumpulan dan pemrosesan data untuk menghasilkan data atau informasi baru yang mudah dipahami dan berguna.

#### Tahapan Proses *Knowledge Discovery in Database* (KDD)

(Bulolo, 2020) menyatakan bahwa proses KDD dapat dipecah menjadi beberapa langkah, seperti yang ditunjukkan pada Gambar 1.1.



**Gambar 1. 1** Tahapan Proses KDD  
Sumber : (Bulolo, 2020)

1. **Data**

Hal pertama yang harus dipersiapkan dalam proses KDD yaitu data.

Data yang di gunakan merupakan data yang sudah terpisah dengan data operasional.

2. *Selection*

Perlu dilakukan pemilihan data karena tidak semua data yang ada dapat dipergunakan. Membuat kumpulan data target, menentukan variabel, memilih sampel data, dan menyimpan data dalam file adalah contoh kegiatan pemilihan data.

3. *Pre-processing (Cleaning)*

Pada titik ini, data yang dipilih akan dibersihkan. Setelah menghapus duplikat dan menyelesaikan inkonsistensi informasi, proses pembersihan selesai. Data juga dapat diperkaya dengan menambahkan informasi tambahan yang relevan, yang dikenal sebagai enrichment, selama tahap praproses ini.

4. *Transformation*

Ada berbagai algoritma dan metode yang dapat digunakan dalam *data mining*. Seperti berdiri, format data yang dibutuhkan oleh setiap algoritma atau metode berbeda. Ketika proses KDD digunakan, data yang telah disiapkan sebelumnya dimasukkan ke dalamnya, dan setiap perubahan yang diperlukan dilakukan terlebih dahulu.

## 5. *Data Mining*

Tahap utama KDD adalah *data mining*. Pengetahuan dan informasi dapat ditambang dari data menggunakan algoritma atau metode tertentu berdasarkan pengetahuan atau informasi yang sedang dicari.

## 6. *Interpretation (Evaluation)*

Proses *data mining* menghasilkan pengetahuan atau informasi yang dapat dengan mudah dipahami oleh pihak terkait, seperti informasi yang ditampilkan dalam bentuk grafik, pohon keputusan, atau aturan. Periksa untuk melihat apakah proses *data mining* menghasilkan pengetahuan atau informasi yang berlainan dengan keyakinan atau fakta yang dimiliki sebelumnya.

## 7. *Knowledge*

Tujuan utama dari proses KDD adalah untuk mengumpulkan pengetahuan atau data yang berguna dan dapat dimengerti. Sesuai dengan manfaat atau kegunaan dari pengetahuan atau informasi yang dihasilkan, itu dilaksanakan.

## 2.2 *Data Mining*

*Data mining* ialah metode yang dipakai buat menciptakan pengetahuan terkini dalam sejumlah besar data. Terdapat beberapa tahap yang wajib diiringi buat melaksanakan prosedur ini (Sitepu and Buulolo, 2017). Tujuan dari *data mining* adalah untuk membuat database dari data yang sebelumnya tidak tersedia. Menganalisis data didefinisikan sebagai proses mencari dan mengidentifikasi koneksi di antara kumpulan data yang berbeda. Data dalam database dapat disaring

secara efisien menggunakan hubungan ini. Ekstraksi informasi terkini dari sejumlah besar data, yang menunjang dalam pengambilan ketentuan, diketahui sebagai *data mining*. Istilah "*knowledge discovery*" kadang-kadang digunakan untuk menggambarkan penambangan data.

Sedangkan menurut (Arhami and Nasir, 2020) *data mining* adalah proses "menggali" melalui data untuk menemukan informasi atau pengetahuan baru yang bermanfaat bagi pengguna. Mengumpulkan, mengekstraksi, menganalisis, dan melaporkan data semuanya termasuk dalam proses *data mining*. Analisis pola data tersembunyi juga dikenal sebagai *data mining*. Analisis gudang data, algoritma *data mining*, dan memfasilitasi pengambilan keputusan bisnis dan informasi lainnya dikumpulkan di area umum dan digunakan untuk tujuan ini.

### **2.2.1 Pengelompokan Data Mining**

Menurut (Buulolo, 2020) *data mining* dibagi menjadi berbagai kategori tergantung pada tugas yang dapat diselesaikan, termasuk:

1. Deskripsi: Guna mempromosikan kegiatan, perlu untuk mengidentifikasi atau menemukan pola yang terjadi secara teratur. Ketika pelanggan sering membeli produk A dan B pada saat yang sama, manajemen supermarket memindahkan A ke lokasi yang berdekatan di katalog. Untuk membuatnya sesederhana mungkin bagi pelanggan untuk membeli produk.

2. **Klasifikasi:** Berdasarkan korelasi antara kondisi dan variabel target. Bencana alam, misalnya, dipecah menjadi tiga kategori: bencana, sedang, dan non-bencana.
3. **Prediksi:** Prediksi dan klasifikasi sangat mirip. Peramalan adalah fungsi penambangan data yang umum. Nilai hasil prediksi akan digunakan di masa depan berdasarkan data sebelumnya.
4. **Estimasi:** Menurut definisi istilah "estimasi," itu adalah prediksi; Namun, berbeda dari klasifikasi dalam bahwa estimasi adalah angka daripada pengelompokan abjad.
5. **Pengklasteran:** Pengklasteran adalah sekelompok data yang memiliki karakteristik umum (tipe yang sama). Pengamatan, catatan data, atau kelas dan objek serupa adalah contoh data yang dapat dikelompokkan ke dalam Pengklasteran.
6. **Asosiasi:** Asosiasi adalah pengelompokan, serikat pekerja, atau jenis pengelompokan lainnya. Contoh umum dari *data mining* adalah mengejar atribut yang muncul atau selalu muncul bersamaan dengan pembelian lebih dari satu item, seperti ketika Anda membeli produk A, Anda juga membeli B, Anda juga membeli B, Anda juga membeli C dan sebagainya.

### 2.2.2 Prediksi

Peramalan adalah kombinasi dari seni dan sains. Prediksi Adalah mungkin untuk memprediksi nilai masa depan dengan berfokus pada data dan informasi yang relevan, apakah itu informasi serta data dari periode lalu ataupun data dari periode saat ini. Prediksi dapat dilakukan dengan berbagai cara, termasuk keduanya:

#### 1. Metode Kualitatif

Model matematika tidak diperlukan untuk menggunakan metode ini. Akibatnya, tidak mungkin untuk memprediksi masa depan hanya berdasarkan data yang dikumpulkan (peramalan jangka panjang). Adalah umum bagi perkiraan kualitatif untuk mengandalkan pendapat para ahli di bidangnya.

Selain itu, metode ini hemat biaya karena tidak memerlukan banyak data dan dapat diperoleh dalam hitungan menit. Namun, kelemahan metode ini adalah sifat subjektif dari data yang dihasilkannya.

#### 2. Metode Kuantitatif

Untuk memprediksi masa depan, metode ini bergantung pada data mentah dan aturan matematis yang menyertainya. Metode kuantitatif memiliki banyak model yang berbeda untuk membuat prediksi, seperti:

##### a. Model-model regresi

Memprediksi variabel yang mempunyai ikatan linier dengan variabel bebas yang dikenal serta bisa diharapkan ialah tujuan dari model ini.

b. Model Ekonometrik

Segmen ekonomi dirangsang oleh variabel independen dalam persamaan regresi yang digunakan dalam model.

c. Model *Time Series Analysis* (Deret Waktu)

Ini adalah model yang menggunakan data dari masa lalu untuk memprediksi tren masa depan berdasarkan data saat ini.

### 2.2.3 *Time Series Analysis*

Istilah "*Time Series*" mengacu pada kumpulan data yang mencakup periode waktu tertentu. Prediksi data masa depan menggunakan persamaan matematika dan statistik juga merupakan bagian dari peramalan deret waktu. Data *Time Series* tersedia dalam berbagai bentuk, termasuk: (Octavia, Tanti and Yulia, Yulia dan Lydia, 2015):

a. Siklus

Pola seperti ini dapat dilihat dengan cara pola ini naik dan turun. Pola periodik dapat ditemukan dengan menghapus pola musiman dari kumpulan data jika digunakan mingguan atau bulanan.

b. *Random*

Pola acak yang tidak dapat digambarkan dengan menggambar. Sebagai akibat dari keadaan darurat, tidak ada cara untuk memprediksi atau menjelaskan ketidakteraturan pola ini.

c. *Trend*

Peningkatan atau penurunan komponen jangka panjang pola data dapat dilihat sebagai tren karena pola tren berubah dari waktu ke waktu.

d. Musiman

Dalam sebuah pola, gerakan terbukti diulang dari waktu ke waktu. Data plakat yang dikumpulkan mingguan dan bulanan dapat menunjukkan tren ini. Metode *moving averages*, *smoothing exponential* musim dingin klasik adalah metode umum untuk menentukan nilai pola musiman.

Teknik peramalan *Time series* terdiri atas:

1. “Statisika”
  - a. *Moving Average*
  - b. *Exponential Smoothing*
  - c. Regresi
  - d. *ARIMA ( Box Jenkins )*
2. Kecerdasan buatan
  - a. *Neural Network*
  - b. Algoritma Genetika
  - c. *Simulated Annealing*
  - d. *Genetic Programming*
  - e. Klasifikasi
  - f. *Hybrid”*



## 2.2 Metode *Data Mining*

Metode klasifikasi digunakan untuk menemukan model yang menggambarkan dan membedakan kelas konseptual data oleh penulis dalam riset ini. Data yang dipakai guna membangun model ini diperoleh lewat pemeriksaan *data set* praktik. Dengan menggunakan model ini, bisa memperkirakan label kelas dari objek yang tak dikenal (Nikmatun and Waspada, 2019).

Menurut (Muslim *et al.*, 2019) Klasifikasi adalah strategi untuk memprediksi kelas entitas yang labelnya tidak diketahui dengan menemukan model yang menjelaskan atau membedakan gagasan atau kelas fakta. Sedangkan menurut (Iriadi & Nuraeni, 2016) dalam (Wijaya and Fauzi, 2020), klasifikasi data ialah proses dalam menemukan suatu properti-properti sama pada himpunan objek pada database, serta mengklasifikasikan ke kelas-kelas berbeda. Tujuannya yaitu menemukan model pada *training set* yang membedakan antara atribut ke dalam kategori atau kelas yang sesuai.

Dari sekian banyak teknik dalam pengklasifikasian, diantaranya yaitu *K-Nearest Neighbour*, *Naïve Bayes*, *Decision Tree*, *Neural Networks-Support Vector Machine* dan lain-lain.

## 2.3 Algoritma *K-Nearest Neighbour* (K-NN)

Sejak tahun 1970, teknik *K-Nearest Neighbour* (K-NN) telah digunakan dalam estimasi statistik dan pengenalan pola (Arhami and Nasir, 2020). K-NN adalah salah satu pendekatan non-parametrik pada saat itu. K-NN adalah algoritma klasifikasi atau pendekatan yang banyak digunakan dalam *data mining*. K-NN juga

termasuk kedalam kategori regresi yang juga dapat digunakan untuk memprediksi seperti halnya regresi.

Secara umum, ada 2 komponen yang paling penting dalam K-NN, yaitu :

1. K sebagai parameter yang akan melingkupi sejauh mana atau sejumlah data mana yang akan menjadi ukuran untuk pertimbangan penentuan label atau kelas dari objek latih.
2. Jarak, jarak antara item data yang akan diuji dan semua objek data pelatihan yang diketahui harus diketahui sehingga penempatan objek pelatihan yang lebih dekat dengan tetangganya dapat ditentukan. Untuk menentukan jarak terdekat tersebut maka dapat dipilih fungsi atau metode jarak mana pun yang sesuai dengan kasus yang akan diselesaikan. Jarak yang berbeda akan menentukan posisi atau letak yang berbeda.

Tahapan metode *K-Nearest Neighbor*, antara lain:

1. Parameter k (jumlah tetangga terdekat) ditentukan.
2. Menghitung jarak antara data yang akan dinilai dan data yang akan dievaluasi dengan pelatihan penuh;
3. Mengurutkan jarak yang dihasilkan;
4. Menentukan jarak yang paling dekat dengan urutan;
5. Menempatkan kelas yang sebanding bersama-sama;
6. Carilah jumlah kelas yang dibagikan oleh tetangga yang berdekatan dan tetapkan kelas-kelas tersebut untuk dinilai sebagai kelas data.

Rumus Euclidean Distance persamaan 2.1 dapat digunakan untuk menghitung jarak antara dua titik, seperti titik data *training* dan titik data *testing*.

Persamaan 2.1 adalah rumus untuk jarak komputasi menggunakan algoritma K-Nearest Neighbor (K-NN) dan *Euclidean Distance*:

$$d(P, Q) = \sqrt{\sum_{i=1}^n (P_i - Q_i)^2} \quad \text{Rumus 2. 1 Perhitungan Jarak Euclidean}$$

Dimana

$d(P, Q)$  : jarak *euclidien*

$n$  : jumlah data *training*

$P$  : inputan data ke -1 dari data *training*

$Q$  : inputan data ke -1 dari data *testing*

Ada sejumlah teknik pembelajaran berbasis kasus yang termasuk dalam pendekatan K-Nearest Neighbor (K-NN) untuk pembelajaran mesin. Algoritma juga merupakan metode pembelajaran yang tidak efisien. Di K-NN, kelompok objek terdekat (serupa) dengan objek ditemukan menggunakan data baru atau data uji yang berasal dari data pelatihan (Islami, 2018). Data pembelajaran yang paling dekat dengan objek digunakan untuk mengklasifikasikannya menggunakan algoritma *K-Nearest Neighbor*. Kasus-kasus baru ditemukan dengan menghitung jarak antara kasus baru dan kasus lama yang cocok dengan sejumlah fungsi yang ada berdasarkan berat kasus baru. Hal ini dimungkinkan untuk menggunakan rumus *Euclidean* untuk mengetahui jarak antara dua titik, yaitu  $x$  dalam data *training* dan  $y$  dalam data *testing*.

D adalah jarak antara titik pada data pelatihan  $x$  dan titik pengujian data  $y$  yang akan dikategorikan, di mana  $x=x_1,x_2,\dots,x_i$  dan  $y=y_1,y_2,\dots,y_i$  dan  $I$  menunjukkan nilai atribut dan  $n$  menunjukkan dimensi atribut.

Tahapan menghitung metode Algoritma *K-Nearest Neighbor*, yakni:

1. Parameter  $K$  (Jumlah tetangga terdekat) sedang ditentukan.
2. Menghitung kuadrat setiap objek dari jarak *euclid (instance query)* menggunakan contoh data yang diberikan;
3. Setelah itu, atur item ke dalam kelompok dengan jarak Euclidean terpendek;
4. Mengumpulkan kategori  $Y$  (Klasifikasi *Nearest Neighbor*);
5. Sebagian besar nilai *instans query* yang telah ditentukan dapat diantisipasi dengan menggunakan kategori *Nearest Neighbor*.

#### 2.4 *Software Pendukung*



**Gambar 2.1** Logo *Rapidminer*  
Sumber : ([www.rapidminer.com](http://www.rapidminer.com), 2021)

Pada penelitian ini peneliti menggunakan *software RapidMiner*. *RapidMiner* merupakan *software data mining* gratis untuk keperluan akademik dimana bertujuan untuk memproses *data mining* dengan cakupan analisis teks, mengestrak pola yang ada dari kumpulan data besar dengan menggunakan metode statistik, kecerdasan buatan, dan database untuk digabungkan. Tujuan analisis ini yakni guna mendapat informasi dengan nilai kriteria tertinggi dari data yang diolah.

## 2.5 Penelitian Terdahulu

Penelitian bergantung pada penelitian terdahulu. Berikut penelitian yang dapat berfungsi sebagai panduan untuk penelitian ini, termasuk:

1. Penelitian oleh (Sitepu and Buulolo, 2017) Nomor 1 Vol.7 dengan judul **“IMPLEMENTASI ALGORITMA NEAREST NEIGHBOR PADA PENERIMAAN PEGAWAI BARU PADA MTS IKHWANUTS TSALITS TALUN KENAS”**, penelitian ini menyatakan bahwa K-NN efektif dalam sistem penerimaan pegawai dengan cara penghitungan nilai kedekatan diantara kasus lama dengan kasus baru. Dikarenakan setiap nilainya dapat ditentukan oleh pengguna (*user*) dengan persyaratan yang di tentukan yaitu nilai IPK, nilai TOEFL, prestasi, pengalaman, umur, dan status. Maka dari itu penelitian ini dapat dijadikan rujukan.
2. Penelitian (Winarso and Arribe, 2017) Nomor 2 Vol.8 dengan judul **“SELEKSI PEGAWAI DAN DOSEN UMRI BERBASIS E-RECRUITMENT MENGGUNAKAN METODE K-NEAREST NEIGHBOR”**, menyatakan bahwa Dalam seleksi administrasi, metode K-NN dapat digunakan untuk memilih calon karyawan dan dosen dengan

menghitung kesamaan antara persyaratan dan data pelamar. Universitas Muhammadiyah Riau juga telah berhasil mengembangkan sistem e-recruitment untuk mengevaluasi calon karyawan dan dosen. Penelitian ini dapat berfungsi sebagai panduan.

3. Penelitian oleh (Khasanah, Harjoko and Candradewi, 2016) Nomor 2 Vol. 6 yang berjudul “**KLASIFIKASI SEL DARAH PUTIH BERDASARKAN CIRI WARNA DAN BENTUK DENGAN METODE *K-NEAREST NEIGHBOR (K-NN)***” menyatakan Mengklasifikasikan sel darah berdasarkan jenisnya menggunakan mikroskop telah lama menjadi praktik standar di laboratorium hematologi. Laboratorium hematologi menggunakannya sebagai alat diagnostik dan pemantauan utama. Ada juga prosedur manual yang memakan waktu untuk lulus serangkaian tes laboratorium. Akibatnya, penelitian ini berfokus pada tahap awal klasifikasi otomatis jenis sel darah putih di bidang medis. Morfologi sel darah dapat digunakan dalam teknik pengolahan gambar untuk mengatasi masalah waktu dan diagnosis dini. Peneliti menggunakan K-Nearest Neighbor untuk mengklasifikasikan sel darah putih berdasarkan morfologi sel (K-NN). Hough circle, threshold, dan feature extraction adalah algoritma pemrosesan gambar yang digunakan. Akhirnya, klasifikasi K-Nearest Neighbor (K-NN) digunakan. Untuk mengidentifikasi jenis gambar, 100 gambar dianalisis. Hasil dari tes segmentasi dan klasifikasi menunjukkan tingkat akurasi masing-masing 78% dan 64%.

4. Penelitian (Lizarti and Ulfah, 2019) Nomor 1 Vol.4 berjudul **“PENERAPAN ALGORITMA *K-NEAREST NEIGHBOR* UNTUK PENENTUAN PEMINATAN STUDI STMIK AMIK RIAU”** menyatakan Mahasiswa di STMIK Amik Riau didorong untuk mengejar minat akademik mereka berdasarkan bakat dan minat mereka yang unik. Mahasiswa dalam Program Studi Teknik Informatika STMIK Amik Riau menghususkan diri dalam bisnis dan jaringan. Kemampuan dan minat siswa harus dipertimbangkan ketika memilih minat. Nilai mahasiswa dan tingkat kelulusan sangat dipengaruhi oleh tingkat minat mereka dalam kursus. Seperti berdiri, pemilihan minat belajar mahasiswa dibuat semata-mata atas dasar persahabatan dan bukan kemampuan akademik. Karena diyakini memberikan saran yang baik dan tepat untuk belajar, itu menjadi jawaban atas masalah memilih jurusan atau konsentrasi studi. Pengelompokan data dapat dilakukan dengan menggunakan algoritma klasifikasi *K-Nearest Neighbor* (K-NN). Data nilai kursus prasyarat dari semester pertama hingga kelima digunakan dalam penelitian ini. Algoritma K-NN dapat diimplementasikan menggunakan aplikasi berbasis PHP dan *MySQL*. Jika dibandingkan dengan hasil yang dihitung secara manual, output sistem 100% akurat. Untuk menguji kinerja algoritma, gunakan alat *RapidMiner*. Algoritma K-NN bekerja dengan baik pada 183 data pelatihan dan 100 data pengujian, dengan *accuracy*, *Recall*, *Precision*, *F Measure*, dan *Classification Error* hasil masing-masing 98%, 100%, 100%, 91,67%, dan 2%. Mahasiswa Teknik Informatika STMIK Amik

Riau dapat memperoleh manfaat dari penelitian ini dengan menerima saran minat studi.

5. Penelitian (Imron and Kusumah, 2018) Nomor 1 Vol. 1 dengan judul ***“APPLICATION OF DATA MINING CLASSIFICATION METHOD FOR STUDENT GRADUATION PREDICTION USING K-NEAREST NEIGHBOR (K-NN) ALGORITHM”*** menyatakan bahwa tingkat kelulusan mahasiswa merupakan salah satu indikator untuk meningkatkan akreditasi suatu mata kuliah. Diperlukan untuk memantau dan mengevaluasi kelulusan siswa kecenderungan, tepat waktu atau tidak. Salah satunya adalah memprediksi tingkat kelulusan dengan memanfaatkan teknik *data mining*. Metode Klasifikasi *Data mining* yang digunakan adalah algoritma *K-Nearest Neighbor* (K-NN). Data yang digunakan berasal dari data siswa, data nilai siswa, dan data kelulusan siswa untuk tahun 2010-2012 dengan total 2.189 catatan. Atribut yang digunakan adalah jenis kelamin, asal sekolah, program studi IP Semester 1-6. Hasil menunjukkan bahwa metode K-NN menghasilkan akurasi yang tinggi sebesar 89,04%.4.



## 2.6 Kerangka Pemikiran

Perlu menggunakan metode K-NN untuk memprediksi perekrutan karyawan di PT. DSAW. Perangkat lunak *RapidMiner* digunakan dalam penelitian ini untuk menganalisis data sebelumnya dan memprediksi efisiensi rekrutmen. Kerangka penelitian didasarkan pada latar belakang dan metode yang digunakan:



**Gambar 2.1** Kerangka pemikiran

Sumber : Data Penelitian ( 2021 )

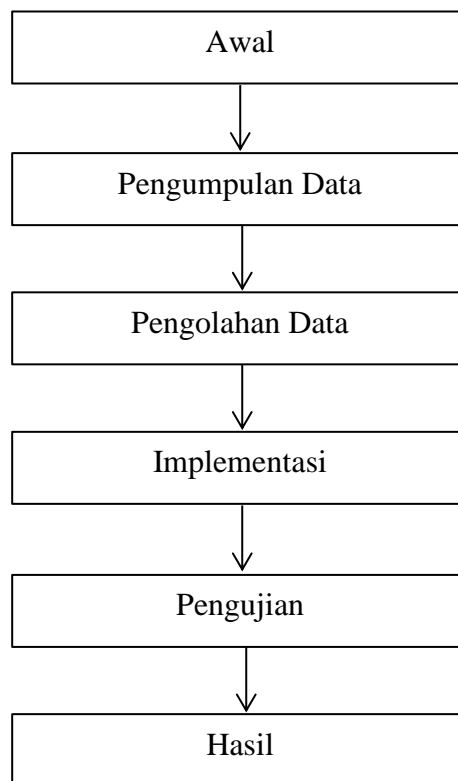


**Universitas Putera Batam**

## **BAB III METODE PENELITIAN**

### **3.1 Desain Penelitian (Arsitektur Sistem)**

Desain penelitian studi ini menggambarkan perkembangan suatu proyek dari tahap awal hingga puncaknya.



**Gambar 3.1** Desain Penelitian

Sumber : Data Penelitian ( 2021 )

Peneliti dapat menjelaskan hal berikut berdasarkan strategi penelitian yang disebutkan di atas:

1. Awal

Pemahaman akan menjadi masalah utama penelitian pada tahap awal proyek ini. Kenali beberapa teori tentang penelitian dan bagaimana menemukan metode dan hasil untuk memecahkan masalah penelitian sehingga dapat mencapai hasil yang diinginkan.

2. Pengumpulan Data

Data yang digunakan dalam penelitian ini sejauh ini berasal dari statistik perekrutan staf di PT. DSAW selama tiga tahun terakhir. Kami akan menggunakan informasi ini sebagai data mentah. Ini akan digeneralisasikan, kemudian diubah menjadi format yang dapat diproses oleh *data mining* untuk digunakan nanti.

3. Pengolahan Data

*RapidMiner* dipakai guna memproses data yang sudah dimasukkan ke dalam aplikasi *RapidMiner* dalam bentuk informasi.

4. Implementasi

Penulis sering menggunakan aplikasi *RapidMiner* pada saat ini untuk membantu mereka membuktikan temuan mereka dalam penelitian mereka. Ini akan digunakan oleh PT. DSAW untuk mengevaluasi efisiensi rekrutmen.

5. Pengujian

*RapidMiner* akan digunakan untuk menguji data yang dikumpulkan, dan hasilnya akan dipakai untuk menarik kesimpulan dari data yang diperiksa pada saat

ini. Pengujian diawali dengan pemasukan data, yang kemudian diolah oleh program *RapidMiner*.

#### 6. Hasil

Pada titik inilah dalam proses penelitian semua data telah diproses dan informasinya memiliki nilai prediksi yang tinggi.

### 3.2 Teknik Pengumpulan Data

Ketika suatu kasus dipecahkan, metode ialah cara atau teknik yang digunakan untuk mengumpulkan data secara sistematis. Akibatnya, penulis menggunakan berbagai metode untuk mendapatkannya, seperti:

#### 1. Studi Lapangan (*Field Research*)

Istilah "penelitian lapangan" mengacu pada praktik memperoleh data langsung dari suatu lokasi atau objek. Penulis menggunakan metode berikut untuk mengumpulkan data:

##### a. Wawancara (*Interview*)

Tanya jawab langsung dengan orang dalam yang relevan adalah metode pengumpulan data melalui wawancara. Seorang karyawan PT. DSAW diwawancarai langsung oleh penulis.

##### b. Pengamatan (*Observation*)

Metode yang sangat efektif untuk memperoleh informasi tentang subjek studi. PT. DSAW diamati langsung oleh penulis.

##### c. Sampel

Jika ingin mengambil sampel atau memberi contoh, maka dapat menggunakan metode pengumpulan data ini. Penelitian terhadap perekrutan

karyawan di PT. DSAW dilakukan oleh penulis. Adapun sample yang akan diambil adalah dari tahun 2018, 2019 dan 2020 dengan departemen seperti tabel berikut:

**Tabel 3. 1** Sampel Data Perekrutan Karyawan

TAH UN	DEPARTE MEN	BULAN											
		JA N	FE B	MA R	APR IL	MA Y	JU N	JU L	AU G	SEP T	OC T	NO V	DE C
2020	<i>Accounting</i>	0	0	0	0	0	0	0	0	0	0	1	0
	<i>Com &amp; Log</i>	0	0	0	0	0	0	0	0	1	0	0	0
	<i>Engineer</i>	0	0	0	0	0	0	0	0	0	1	0	0
	<i>HRD &amp; GA</i>	0	0	0	0	1	1	1	1	1	0	0	0
	<i>HSE</i>	0	0	0	2	0	0	0	0	0	0	0	0
	<i>MTC</i>	0	0	0	0	0	0	0	3	4	0	0	1
	<i>PPC</i>	0	0	0	0	0	0	0	1	0	0	0	0
	<i>PROD</i>	1	5	2	2	0	0	0	0	57	3	10	10
	<i>QA</i>	0	0	0	0	0	0	0	0	1	0	0	1
	<i>QC</i>	1	0	1	1	1	0	0	0	3	3	1	1
	<i>STORE</i>	0	0	2	0	0	0	0	0	0	1	0	0
<i>WE</i>	0	0	1	0	0	0	0	0	0	1	0	0	
TOTAL		2	5	6	5	2	1	1	5	67	9	12	13

DEPARTEMEN		JA N	FE B	MA R	APR IL	MA Y	JU N	JU L	AU G	SEP T	OC T	NO V	DE C
2019	<i>Accounting</i>	0	0	0	0	0	0	0	0	0	0	1	0
	<i>Com &amp; Log</i>	0	0	0	0	0	0	0	0	0	0	0	0
	<i>Engineer</i>	0	0	0	0	0	0	0	0	0	0	0	0
	<i>HRD &amp; GA</i>	0	0	1	0	0	0	0	0	0	1	0	0
	<i>HSE</i>	0	0	0	3	0	0	0	0	0	0	0	0
	<i>MTC</i>	3	0	3	0	0	0	0	0	2	2	0	0
	<i>PPC</i>	0	0	1	0	1	0	0	0	1	0	0	0
	<i>PROD</i>	34	6	7	8	0	0	3	36	33	10	0	8
	<i>QA</i>	0	0	0	0	0	0	0	0	0	0	0	0
	<i>QC</i>	0	0	0	0	0	0	0	0	3	0	0	4
	<i>STORE</i>	0	0	2	0	0	0	0	0	0	1	0	0
<i>WE</i>	0	0	0	0	0	0	0	0	5	1	2	0	
TOTAL		37	6	14	11	1	0	3	36	44	15	3	12

DEPARTEMEN	JAN	FEB	MAR	APRIL	MAY	JUN	JUL	AUG	SEPT	OCT	NOV	DEC
2018	<i>Accounting</i>	0	0	0	0	0	0	0	4	0	0	0
	<i>Com &amp; Log</i>	0	0	0	0	0	0	0	11	0	0	0
	<i>Engineer</i>	0	0	0	0	0	0	0	2	0	0	0
	<i>HRD &amp; GA</i>	0	0	0	0	0	0	0	3	0	0	0
	<i>HSE</i>	0	0	0	0	0	0	0	4	0	0	0
	<i>MTC</i>	0	0	0	0	0	0	0	9	0	0	0
	<i>PPC</i>	0	0	0	0	0	0	0	3	0	0	0
	<i>PROD</i>	0	0	0	0	0	0	0	68	0	0	0
	<i>QA</i>	0	0	0	0	0	0	0	0	0	0	0
	<i>QC</i>	0	0	0	0	0	0	0	7	0	0	0
	<i>STORE</i>	0	0	0	0	0	0	0	3	0	0	0
<i>WE</i>	0	0	0	0	0	0	0	28	0	0	0	
TOTAL	0	0	0	0	0	0	0	0	142	0	0	0

Sumber : Data Penelitian ( 2021 )

d. Studi Kepustakaan (*Library Research*)

Penulis melakukan studi pustaka dan mengumpulkan informasi tentang penulisan makalah dari berbagai sumber, diantaranya buku, jurnal, dan internet.

2. Prosedur Rancangan

Dokumen seperti buku dari perpustakaan, modul, dan manual, serta yang lain yang dianggap perlu atau berguna, digunakan dalam proses ini untuk mengumpulkan data.

### 3.3 Operasional Variabel

Peneliti dapat mengumpulkan data dan menarik kesimpulan tentang berbagai hal dengan memanipulasi variabel. Variabel-variabel dalam penelitian ini diinterpretasikan sebagai berikut:

1. Jumlah calon karyawan

Angka ini akan menjadi tolak ukur keberhasilan atau pencapaian. Efisiensi rekrutmen dapat diprediksi dengan menggunakan proses rekrutmen bulanan. Penelitian ini sangat dipengaruhi oleh banyaknya orang yang tertarik untuk bergabung dengan perusahaan tersebut.

### 3.4 Metode Analisis dan Rancangan Sistem ( *Data Mining* )

Desain sistem aplikasi *data mining* menggunakan metode k-terdekat membutuhkan analisis dan desain sistem. Dalam analisis kebutuhan, persyaratan sistem ditata. Untuk merancang sistem, penting untuk mempertimbangkan hal-hal berikut:

1. Input

Rekapitulasi transaksi dari 3 tahun terakhir digunakan sebagai masukan dalam pembangunan sistem *data mining* untuk prediksi perekrutan karyawan menggunakan metode *K-Nearest Neighbour*.

2. Proses

Peneliti memanfaatkan implementasi dengan *K-Nearest Neighbour* untuk menentukan sistem atau prosedur mana yang akan lebih berhasil dan efisien dalam memutuskan kandidat potensial untuk diterima sebagai karyawan organisasi.

3. Output

Output atau hasil yang dapat disimpulkan dalam penelitian ini adalah untuk mengetahui tingkat efisiensi dan efektifitas proses rekrutment.



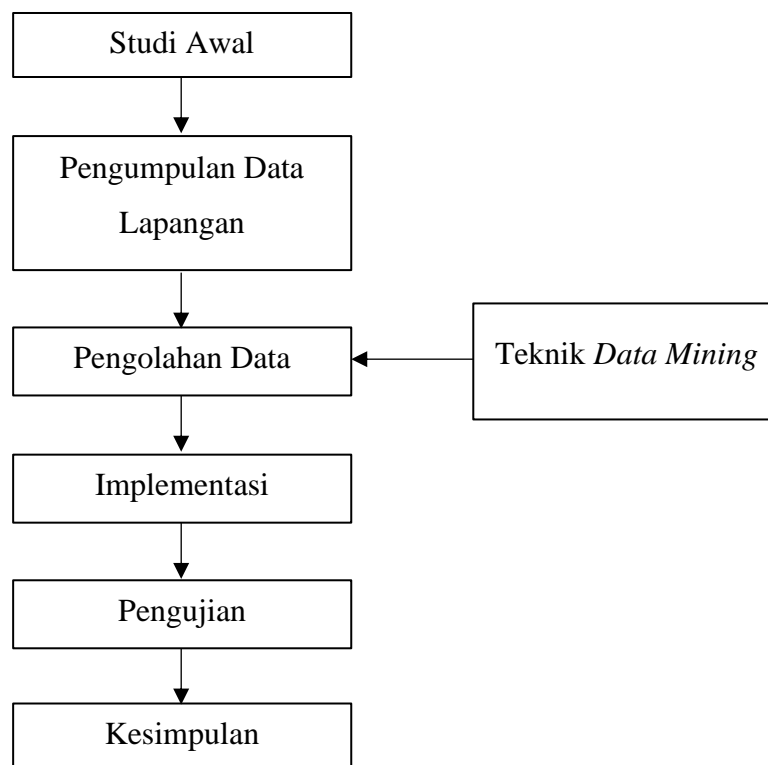
Adapun kebutuhan pendukung dalam proses perancangan sistem tersebut diantaranya:

A. *Hardware*

Berikut ini adalah spesifikasi perangkat keras yang digunakan dalam penelitian ini: Laptop

B. *Software*

Berikut ini adalah langkah dari proses pengolahan data karyawan di PT. DSAW menggunakan algoritma *K-Nearest Neighbour* dimana menggunakan perangkat lunak olah data pada penelitian ini yaitu: *RapidMiner*.



**Gambar 3.2** *Flowchart Of Document Analisis Kebutuhan Sistem*

Sumber : Data Penelitian ( 2021 )

Langkah-langkah diagram alur penelitian yang dibuat adalah sebagai berikut:

1. Studi Awal

Langkah pertama yang penting adalah mengidentifikasi dan menyelidiki masalah penelitian. Maka perlu untuk mempelajari berbagai studi literatur yang berkaitan dengan masalah khusus ini dan menentukan bagaimana memecahkan masalah sehingga Anda dapat memenuhi tujuan yang Anda nyatakan. Studi ini sangat bergantung pada literatur. Langkah selanjutnya adalah memilih literatur yang telah dipelajari sebagai dasar penelitian. Tinjauan literatur teori penambangan data dan metode autoregressive integral dari rata-rata bergerak disajikan di sini.

2. Pengumpulan Data

Sebagai bagian dari pengumpulan data riset ini, tata cara pencatatan file data perekrutan karyawan dipakai sebagai dasar buat penerapan serta percobaan.

3. Pengolahan Data

Perusahaan sering mengalami masalah identifikasi pada saat ini dalam proses pengolahan data. Kami akan berbicara tentang PT. DSAW dan isu-isu yang muncul sehingga kami dapat menemukan solusi. Setelah masalah dianalisis, langkah selanjutnya adalah bagi penulis untuk menentukan bentuk solusi dan pemahaman yang kuat tentang bagaimana cara menyelesaikannya. Metode built-in rata-rata pergerakan autoregressive digunakan oleh penulis dalam penambangan data untuk menghasilkan hasil yang sejalan dengan tujuan PT. DSAW.

#### 4. Implementasi

Untuk saat ini, penulis mengandalkan alat dan aplikasi penambangan data yang sudah ada sebelumnya untuk membantu dalam proses memverifikasi data temuan. Data disimpan dalam database, tetapi peneliti lebih suka bekerja di Microsoft Excel.

#### 5. Pengujian

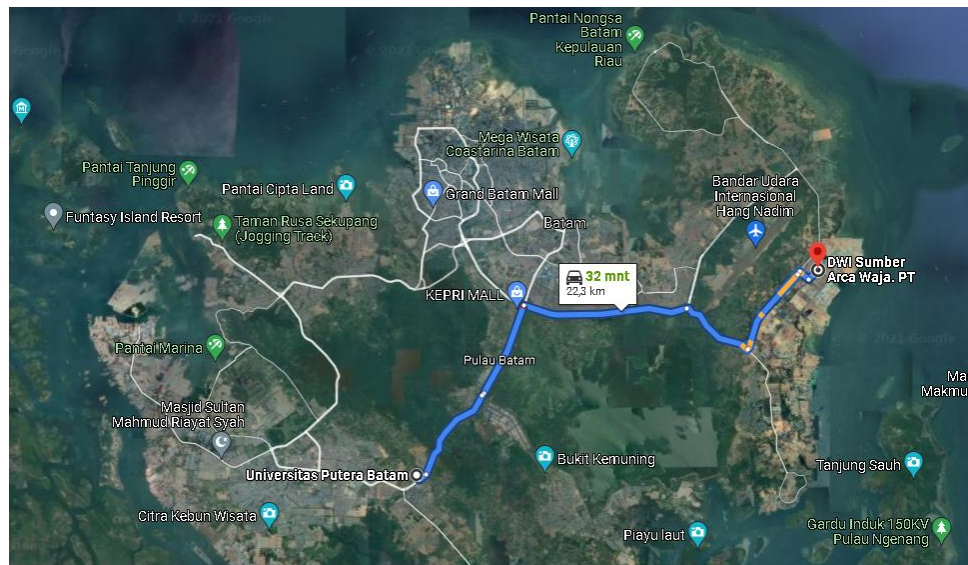
Aplikasi *data mining* digunakan untuk terhubung ke sistem yang telah diisi atau berisi data yang akan diuji pada tahap ini. Persiapan perangkat keras dan perangkat lunak adalah langkah pertama dalam mekanisme pengujian sistem. Aplikasi penambangan data yang dapat dijalankan di komputer tersedia.

### 3.5 Lokasi Dan Jadwal Penelitian

Berikut ini adalah deskripsi jadwal dan proses penelitian lokasi:

#### 3.5.1 Lokasi Penelitian

Penelitian dijalankan di PT. DSAW, yang mana PT. DSAW merupakan perusahaan pemasok migas yang bergerak di bidang *Manufacturing of Large Steel Pipes* dengan tujuan memenuhi permintaan sektor minyak dan gas bumi yang bertempat di Jl. Hang Kesturi I No.8 Kawasan Industri Kabil Batam, 447R+M8H, Batu Besar, Nongsa, Kota Batam, Kepulauan Riau 29467.



**Gambar 3. 1** Lokasi Penelitian  
Sumber : ([www.google.com](http://www.google.com), 2021)

### 3.5.1 Jadwal Penelitian

Program penelitian ini dilaksanakan antara September 2021 dan Januari 2022, dengan kerangka waktu berikut:

**Tabel 3. 2** Jadwal Penelitian

No.	Kegiatan	September				Oktober				November				Desember				Januari 2022			
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
1	Input Judul	■	■	■	■																
2	BAB 1					■	■	■	■												
3	BAB 2									■	■	■	■								
4	BAB 3													■	■	■	■				
5	BAB 4													■	■	■	■	■	■	■	■
6	BAB 5																	■	■	■	■
7	Penyelesaian																				■

Sumber : Data Penelitian ( 2021 )



**Universitas Putera Batam**