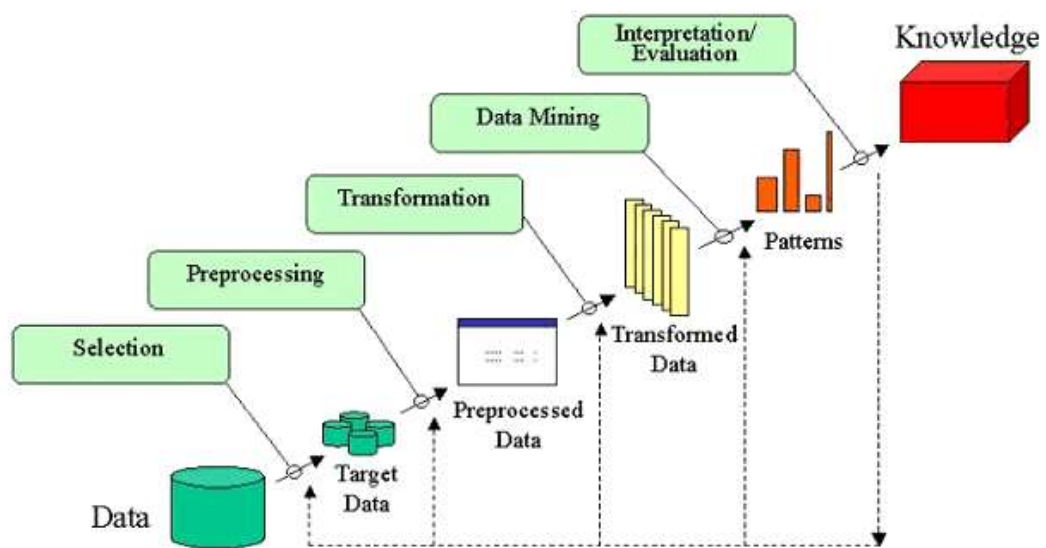


BAB II

TINJAUAN PUSTAKA

2.1 Knowledge Discovery in Database (KDD)

Proses penemuan pengetahuan dari kumpulan data juga sering disebut sebagai *knowledge discovery in database* (KDD) (Elisa, 2018). Semua tahapan dari pengumpulan dan penemuan data dalam skala besar yang didasari oleh pengetahuan dan informasi merupakan KDD. Metode KDD ini mampu memberikan pengetahuan dan informasi yang berguna serta bisa dipahami (Panggabean et al., 2020). Adapun tahapan proses *knowledge discovery in database* (KDD) yaitu sebagai berikut.



Gambar 2. 1 Tahapan Proses KDD

Sumber: (Rizky & Ardian, 2019)

1. Data

Persiapan data diperlukan sebelum memulai proses KDD. Data operasional yang terpisah akan digunakan untuk tujuan ini.

2. *Selection*

Pemilihan data diperlukan karena tidak semua data yang tersedia dapat digunakan dalam proyek tertentu. Salah satu cara untuk memikirkannya adalah bahwa proses ini mengambil pengumpulan data tertentu sebagai titik awal untuk pemrosesan berikutnya.

3. *Pre-processing/cleaning*

Ditahapan ini, data yang berlebihan, tidak konsisten, dan data yang tidak sesuai diperbaiki. Dalam proses preprocessing/cleaning, pengayaan adalah proses penambahan informasi tambahan yang berguna dan dapat diandalkan ke dalam data.

4. *Transformation*

Pola pencarian basis data dan jenis informasi yang akan dicari merupakan pertimbangan penting dalam proses ini.

5. Data mining

Selama tahap proses KDD ini, data yang dipilih akan digunakan untuk mencari pola atau informasi yang menarik. Menurut tujuan atau informasi yang dicari dalam data mining ada beberapa metode, strategi dan algoritma yang berbeda yang dapat digunakan.

6. *Interpretation/Evaluation*

Hasil data mining harus dikomunikasikan dengan cara yang dapat dimengerti oleh mereka yang membutuhkannya. Pohon keputusan, grafik, dan aturan semuanya dapat digunakan untuk menyampaikan informasi ini. Bagian dari proses KDD ini terdiri dari pemeriksaan untuk melihat apakah pola atau data yang diperoleh didasarkan pada fakta atau asumsi yang telah diidentifikasi.

7. *Knowledge*

Data mining adalah tentang menemukan informasi relevan yang dapat dipahami oleh banyak orang, dan informasi ini perlu dipraktikkan sesuai dengan nilainya.

2.2 Data Mining

Dalam data mining, gudang basis data yang luas dijelajahi untuk mendapatkan informasi yang dapat digunakan untuk beberapa tujuan. Mengekstrak informasi baru dari sejumlah besar data alternatif dapat digambarkan sebagai data mining. Penemuan pengetahuan kadang-kadang digunakan untuk menggambarkan data mining. Pola atau model baru dapat ditemukan dalam basis data yang sangat besar dengan penambahan data, proses berulang dan partisipatif (basis data masif) (Gunawan, 2019). Mengekstrak pola berharga atau menarik dari data database memungkinkan informasi baru dan lebih berguna untuk dikumpulkan melalui manipulasi data (Syahputra et al., 2018).

Untuk berbagai alasan, bisnis dan organisasi saat ini tidak mampu menyimpan sejumlah besar data dalam silo tanpa menerapkan teknik penambangan data kepada mereka. Efek baik lainnya dari pertumbuhan eksplosif Internet adalah bahwa data sekarang mudah diakses melalui berbagai perangkat keras dan perangkat lunak berkinerja tinggi. Pada saat yang sama, peningkatan tekanan persaingan dari lingkungan luar memerlukan pendekatan yang berbeda dengan mengekstraksi informasi yang terkandung dalam data yang dikendalikan oleh perusahaan atau organisasi untuk meningkatkan pangsa pasar dan pendapatan(Widaningsih, 2019:17).

2.3 Metode Data Mining

Yang dimaksud dengan metode data mining adalah metode aplikasi yang harus disesuaikan dengan tujuan pengguna. Teknik data mining dapat dipecah menjadi kategori berikut.

1. Classification

Klasifikasi adalah proses memeriksa objek data dan mengklasifikasikannya menurut seperangkat kriteria yang telah ditentukan. Untuk mengkategorikan data baru, classifier membangun model berdasarkan data pelatihan sebelumnya. Pelatihan/pembelajaran pada fungsi tujuan yang memetakan setiap set kualitas (fitur) ke satu atau lebih label kelas yang mungkin dapat dicirikan sebagai klasifikasi. Tidak ada jaminan bahwa sistem klasifikasi akan secara akurat mengkategorikan semua dataset, oleh karena itu perlu dilakukan pengukuran kinerja sistem klasifikasi juga (Utomo & Mesran, 2020).

Klasifikasi data mining adalah proses penentuan konsep kesamaan antar fitur dalam suatu kelompok atau kelas (class). Ini adalah salah satu teknik data mining yang paling sering digunakan. Dengan menggunakan metode ini diharapkan dapat ditentukan kelas objek yang tidak diketahui.

2. *Association*

Dalam penambangan data, salah satu caranya adalah dengan menggunakan analisis asosiasi atau penambangan data asosiasi untuk menemukan aturan yang mengatur asosiasi antara kumpulan objek atau karakteristik. Kumpulan data besar dapat mengambil manfaat dari penggunaan aturan asosiasi untuk mengungkap koneksi tersembunyi. Aturan asosiatif atau set aturan digunakan untuk menggambarkan hubungan antara hal-hal yang ditemukan (Wibowo & Jananto, 2020).

3. *Clustering*

Teknik data mining clustering mencoba untuk menemukan kelompok item dengan karakteristik serupa yang dapat diklasifikasikan dari kumpulan objek lain, dengan objek dalam kelompok yang sama menjadi lebih homogen daripada objek dalam kelompok yang berbeda. Indeks Davies-Bouldin digunakan untuk menetapkan evaluasi klaster (DBI). Dengan menggunakan indeks Davies-Bouldin, metode pengelompokan dievaluasi berdasarkan nilai kohesi dan separasi pada metode pengelompokan. Kohesi dalam clustering didefinisikan sebagai sejauh mana data mengikuti centroid cluster. Jarak antar centroid cluster digunakan untuk menghitung pemisahan (Nabila et al., 2021).

4. *Regression*

Sebagai aturan umum, regresi digunakan untuk menentukan mean aritmatika atau mean dari variabel dependen (dijelaskan atau dependen) dengan memeriksa hubungannya dengan satu atau lebih variabel independen (dijelaskan atau independen). Nilai tetap atau diketahui dari variabel penjelas dapat digunakan untuk memprediksi variabel dependen (re-sampling). Ada banyak jenis regresi linier, tetapi yang paling umum adalah regresi linier. Tren permintaan masa depan dapat diramalkan dengan menggunakan teknik regresi linier ini (Hardianti & Sindar, 2021).

5. *Forecasting*

Pendekatan yang disebut Forecasting data mining berusaha meramalkan nilai yang akan dihasilkan dari waktu ke waktu. Data periode sebelumnya dan nilai noise digunakan untuk membuat proyeksi ini.

6. *Sequencing*

Sequence mengacu pada serangkaian tindakan. Adalah tugas metode analisis urutan untuk mencari pola dalam rangkaian peristiwa atau urutan, yang merupakan fungsi yang sangat jelas. Makan adalah contoh dasar sehari-hari. Piring, garpu, sendok, dan lauk pauk biasanya disajikan dalam urutan ini. Penyortiran adalah metode umum yang digunakan oleh pola seperti ini untuk menemukan informasi.

7. *Descriptive*

Dengan penggunaan teknik data mining ini, kami berharap dapat memperoleh pemahaman yang lebih komprehensif tentang temuan tersebut. Akibatnya, perilaku data dapat didekodekan.

2.4 Analisis Regresi Linear Berganda

Analisis hubungan antara banyak variabel bebas dan variabel respon (variabel terikat) dengan menggunakan regresi linier berganda dikenal sebagai regresi linier berganda (Ervan Triyanto, Heri Sismoro, 2019). Untuk keperluan analisis regresi linier berganda, hubungan antara dua variabel atau lebih dapat diukur (Afkarina et al., 2019). Untuk melakukan regresi linier berganda, model persamaannya adalah:

$$Y = a + b_1X_1 + b_2X_2 + b_3X_3 + \dots + b_nX_n$$

Rumus 2. 1 Rumus Regresi
Linier Berganda

Keterangan:

Y : Variabel Dependen atau Terikat

X (1,2,3,...) : Variabel Independen atau Bebas

a : Nilai konstanta

b (1,2,3,...) : Nilai koefisien regresi

Pengujian hipotesis perlu dilakukan terlebih dahulu sebelum melanjutkan ke analisis regresi. Normalitas, homoskedastisitas, autokorelasi, multikolinearitas, dan linieritas adalah beberapa prasyarat untuk analisis regresi.

2.4.1 Uji Normalitas

Untuk menentukan apakah residual terdistribusi secara teratur, asumsi normalitas diterapkan. Estimasi OLS tidak dapat digunakan jika kondisi normalitas tidak terpenuhi. Anderson Darling, Kolmogorov-Smirnov, Jarque-Bera, dan SkewnesKurtosis adalah beberapa tes yang dapat digunakan untuk mengasumsikan distribusi normal.

Tujuan uji normalitas adalah untuk menentukan apakah variabel dependen dan independen model regresi berdistribusi normal, dan jika demikian, apakah nilai residual model juga berdistribusi normal. Nilai residu yang normal atau hampir normal mencirikan model regresi yang berhasil. Ketika uji normalitas Kolmogorov-Smirnov signifikan 0,05, data dianggap abnormal; ketika uji signifikan $> 0,05$, data dianggap normal (Ningsih & Dukalang, 2019).

2.4.2 Uji Homoskedastisitas

Asumsi homoskedastisitas pada nilai residual dari fungsi regresi populasi sangat penting dalam model regresi linier. Plot nilai kuadrat sisa mungkin mengungkapkan penyimpangan dari asumsi homoskedastisitas. Heteroskedastisitas terjadi jika nilai kuadrat residual mengikuti pola yang dapat diprediksi.

Uji heteroskedastisitas merupakan bagian dari uji hipotesis klasik yang kelima dan terakhir. Berdasarkan hasil uji heteroskedastisitas regresi, varians dari residual berbeda dari satu pengamatan ke pengamatan berikutnya. Asumsi kunci dalam regresi adalah bahwa varians dari residual dari satu pengamatan ke pengamatan lainnya tidak mengikuti pola tertentu. Nilai yang tidak merata antara dua varian residual menggambarkan pola asimetris ini. Gejala heteroskedastis menggambarkan gejala yang tidak merata (Nugraha et al., 2021).

2.4.3 Uji Multikolinieritas

Korelasi berganda antar variabel bebas dapat diuji dengan menggunakan uji multikolinieritas. Seharusnya tidak ada korelasi antara variabel independen dalam

model regresi yang layak. Multikolinearitas dalam regresi dapat diidentifikasi dengan melihat nilai VIF dan toleransi. Sedangkan tidak ada multikolinearitas dengan VIF 10 dan toleransi 0,01 jika VIF > 10 dan toleransi > 0,01 maka terjadi multikolinearitas (Ningsih & Dukalang, 2019).

2.4.4 Uji Validitas

Uji analisis kuantitatif pertama meliputi uji keefektifan. Variabel dan indikator penelitian akan diuji validitasnya untuk memastikan keakuratannya. Untuk menunjukkan keakuratan alat ukur dalam menangkap objek yang diukur, uji validitas ini telah dirancang (Nugraha et al., 2021).

2.4.5 Uji Reliabilitas

Keakuratan atau presisi suatu alat ukur inilah yang dimaksud dengan reliabilitas. Rumus Cronbach Alpha digunakan dalam uji ketergantungan (Puspa et al., 2021):

$$A = \frac{K \cdot r}{1 + K - 1) \cdot r} \quad \text{Rumus 2. 2 Rumus Cronbach's Alpha}$$

Ket:

A : Koefisien reliabilitas

K : Jumlah item reliabilitas

r : Rata-rata korelasi antar item

Pengambilan keputusan pada uji reliabilitas adalah sebagai berikut (Umar, 2007):

- a. Jika koefisien reliabilitas ≥ 0.6 maka item memiliki reliabilitas yang baik
- b. Jika koefisien reliabilitas < 0.6 maka item memiliki reliabilitas yang kurang baik

2.4.6 Uji Heteroskedastisitas

Uji heteroskedastisitas bertujuan untuk mengetahui apakah suatu model regresi memiliki pertidaksamaan residual dari satu pengamatan ke pengamatan berikutnya. Istilah homoskedastis digunakan jika residual dari satu pengamatan ke pengamatan lain adalah sama, dan heteroskedastis digunakan jika berbeda. Model regresi dengan dan tanpa heteroskedastisitas keduanya merupakan pilihan yang baik. Uji glejser dapat digunakan untuk mengetahui ada tidaknya heteroskedastisitas. Heteroskedastisitas tidak dapat dideteksi jika $\text{sig} > 0,05$; karenanya, model yang tidak menunjukkan heteroskedastisitas dianggap baik (Ningsih & Dukalang, 2019).

2.5 *Software* Pendukung

Pada penelitian ini, adapun aplikasi yang akan kita manfaatkan yaitu *Microsoft Excel* dan *Software SPSS* untuk melakukan pengujian data.

2.5.1 Microsoft Excel



Gambar 2. 2 Logo *Microsoft Office Excel 2013*

Sumber: (Peneliti,2022)

Microsoft Excel atau *Microsoft Office Excel* adalah sebuah program aplikasi lembar kerja *SpreadSheet* yang dibuat dan didistribusikan oleh *Microsoft Corporation* untuk sistem operasi *Microsoft Windows* dan *Mac OS*. Strategi *marketing Microsoft* yang baik membuat *Microsoft Excel* sebagai salah satu program komputer yang populer digunakan di dalam komputer hingga saat ini (Novita & Wulanditya, 2020).

Pengolahan data pada penelitian ini akan menggunakan *Microsoft Excel*, sebagai lokasi data karyawan dan penilaian dari beberapa indikator variabel yang digunakan. Dalam penelitian ini, *software Microsoft Excel* digunakan sebagai pengolah data awal yang dilakukan mulai dari pembersihan data seperti menghapus data duplikat, melakukan pengecekan data yang kurang sesuai dan lain sebagainya sehingga data siap digunakan dan diolah di *software SPSS*.

2.5.2 Software SPSS



Gambar 2. 3 Logo Software SPSS 25

Sumber: (Peneliti, 2022)

Aplikasi *SPSS (Statistical Product and Service Solution)* adalah salah satu program atau aplikasi di bidang statistik yang mampu memproses data statistik secara cepat dan akurat. Aplikasi SPSS menjadi sangat populer karena memiliki bentuk pemaparan yang baik (memiliki grafik dan tabel), bersifat dinamis (mudah dilakukan perubahan data), dan mudah dihubungkan dengan aplikasi lain dalam preparasi data sampel analisis (Irawan et al., 2021).

SPSS adalah aplikasi yang digunakan untuk melakukan analisis statistika tingkat lanjut, analisis data dengan algoritma machine learning, analisis string, serta analisis big data yang dapat diintegrasikan untuk membangun platform data analisis (Zein et al., 2019).

Pengolahan data pada aplikasi SPSS dilakukan sesuai dengan modul-modul yang tersedia, diantaranya yaitu *data editor, viewer, multidimensional pivot tables, high-resolution graphics, database access, data transformations, electronic distribution, online help*, akses data tanpa tempat penyimpanan sementara, *interface*

dengan *database relational*, *analisis distribusi*, *multiple Session*, *mapping*, *visualization*, dan lain-lain.

2.6 Penelitian Terdahulu

Penelitian data *mining* ini tentunya menggunakan referensi dari beberapa penelitian yang telah dilakukan oleh peneliti sebelumnya. Adapun referensi jurnal sebagai berikut:

1. Penelitian oleh (Alita et al., 2021) dari Universitas Teknokrat Indonesia, E-ISSN: 2460-7258 Vol.15 No.3, dengan judul “**Analisis Uji Asumsi Klasik Dan Uji Koefisien Regresi Linier Berganda Untuk Rekomendasi Jabatan Struktural Pegawai**”. Tujuan dari penelitian ini yaitu memanfaatkan data mining menggunakan metode Regresi Linier Berganda memberikan hasil prediksi rekomendasi pegawai yang berhak untuk menempati jabatan pada instansi tersebut. Implementasi penggunaan metode Regresi Linier Berganda didapatkan dari 40 pegawai terdapat 26 pegawai yang layak diberikan rekomendasi promosi jabatan. Hasil ini menjelaskan metode regresi linier memiliki kinerja yang baik dengan tingkat *error* yang kecil pada pengujian evaluasi performa regresinya. Sedangkan hasil analisis uji asumsi klasik bertujuan untuk menyimpulkan bahwa data terdistribusi secara normal dan regresi dapat terpenuhi dengan baik dengan melakukan uji multikolinieritas, uji heteroskedastisitas dan uji normalitas. Sedangkan untuk analisis uji koefisien regresi linier berganda yang dilakukan

dengan uji T dan uji F yaitu menghasilkan pengujian yang mempunyai pengaruh yang signifikan terhadap variabel kinerja pegawai.

2. Penelitian oleh (Rizky & Ardian, 2019) dari Universitas Pembangunan Panca Budi Medan, E-ISSN: 2321-1784 Vol. 07 No. 08, dengan judul **“Meningkatkan Kinerja Karyawan Untuk Meningkatkan Motivasi Kerja Universitas Pembangunan Panca Budi Medan”**. Penelitian ini membahas tentang pengaruh penurunan level pegawai kinerja karena dampak ketidakpuasan terhadap hasil promosi menggunakan kompetensi tes. Tujuan penelitian ini untuk melihat pengaruh promosi jabatan dengan indikator kedisiplinan dan kerjasama, terhadap kinerja pegawai. dan pengaruh tidak langsung kompetensi terhadap disiplin dan kerjasama terhadap kinerja karyawan. Desain yang digunakan dalam penelitian ini adalah kuantitatif dengan mengambil karyawan yang bekerja di Universitas Pembangunan Pancabudi sebagai populasi.
3. Penelitian oleh (Purwadi et al., 2019) dari STMIK Triguna Dharma E-ISSN : 2615-3475 Vol.18, No.1, dengan judul **"Penerapan Data Mining Untuk Mengestimasi Laju Pertumbuhan Penduduk Menggunakan Metode Regresi Linier Berganda Pada BPS Deli Serdang"**. Dalam penelitian ini membahas teknik estimasi untuk menemukan pola yang terjadi pada data penduduk terutama hal yang berkaitan dengan laju pertumbuhan penduduk pada Badan Pusat Statistik Kabupaten Deli Serdang. Teknik estimasi yang akan digunakan adalah Regresi Linear Berganda. Metode ini dipilih karena

mampu membuat suatu estimasi/prediksi dengan memanfaatkan data-data lama mengenai laju pertumbuhan penduduk. Sehingga dapat dihasilkan suatu pola hubungan antara atribut-atribut yang mempengaruhi laju pertumbuhan penduduk. Metode Regresi Linear Berganda ini bertujuan untuk membuat persamaan regresi dan prediksi terbaik berdasarkan atribut-atribut yang ada.

4. Penelitian oleh (Panggabean et al., 2020) dari STMIK Budi Darma, E-ISSN: 2715-7393 Vol. 7 No. 1, dengan judul **“Penerapan Data Mining Untuk Memprediksi Pemesanan Bibit Pohon Dengan Regresi Linear Berganda”**. Kurangnya pemanfaatan sistem untuk memprediksi estimasi pemesanan bibit pohon disetiap tahunnya merupakan permasalahan yang terdapat dalam instansi ini sehingga diperlukan suatu metode yaitu Algoritma Regresi Linear Berganda. Maka dengan ini dibuatlah Penerapan Data Mining Untuk Memprediksi Pemesanan Bibit Pohon Dengan Regresi Linear Berganda. Algoritma Regresi Linear Berganda yang merupakan metode yang mendukung dalam memperkirakan atau memprediksi target pemesanan untuk periode yang akan datang. Pengujian Algoritma dilakukan dengan menggunakan software SPSS. Dari hasil penelitian yang telah dilakukan maka dapat membantu pihak BPDASHL untuk mempermudah dalam memprediksi pemesanan bibit dengan menggunakan Software SPSS.

5. Penelitian oleh (Rosyadi et al., 2021) dari Universitas Muhammadiyah Pekajangan Pekalongan, E-ISSN: 2477-3042 Vol. 9 No. 1, dengan judul **“Pengaruh Kepuasan Kerja terhadap Kinerja Karyawan RTO dengan *Structural Equation Modeling* (SEM) dan *Partial Least Square* (PLS)”**. Penelitian ini bertujuan untuk mengetahui apa saja yang mempengaruhi kinerja karyawan. Dengan menggunakan metode *Job Description Index* (JDI) kinerja karyawan akan dipengaruhi oleh kepuasan kerja, motivasi kerja dan komitmen organisasi. Data yang digunakan yaitu data karyawan dari perusahaan dan data tersebut diestimasi menggunakan metode *Structural Equation Modeling* (SEM) dan *Partial Least Square* (PLS). Hasil yang didapatkan setelah data dianalisis terlihat bahwa kepuasan kerja memberikan pengaruh kepada motivasi kerja sebesar 43,1%, motivasi kerja memberikan pengaruh kepada komitmen organisasi sebesar 25,6%, komitmen organisasi memberikan pengaruh kepada kepuasan kerja sebesar 52,4%, kepuasan kerja, motivasi kerja, dan komitmen organisasi memberikan pengaruh kepada kinerja karyawan sebesar 78,7%. Sedangkan pada uji signifikan semua hubungan antar konstruk signifikan dan mempunyai hubungan yang positif hanya pada hubungan motivasi kerja terhadap kinerja karyawan saja yang tidak signifikan.
6. Penelitian oleh (Hardianti & Sindar, 2021) dari STMIK Pelita Nusantara, E-ISSN: 2620-4118 Vol. 4 No. 2, dengan judul **“Implementasi Metode Regresi Linier Berganda Untuk**

Memprediksi Kinerja Karyawan Di PT. Timbang Deli Berbasis Web". Prediksi kinerja karyawan bertujuan meningkatkan efektifitas pembagian tugas yang mempengaruhi reward karyawan. Tahapan analisis sistem yang dibangun adalah tahap pertama analisis untuk pendefinisian kebutuhan. Tahap kedua perancangan antar muka grafis untuk tampilan dari masing-masing kebutuhan. Tahap ketiga, semua fungsi pendukung aplikasi seperti hak akses, input data karyawan, dan antar muka grafis mengalami pengintegrasian ke dalam sebuah rangkaian kode program yang utuh. Sistem prediksi dibangun berbasis web menggunakan metode Regresi Linier Berganda dengan variabel-variabel yang ada.

7. Penelitian oleh (Saputra & Fauzi, 2022) dari Universitas Putera Batam, E-ISSN: 2715-6265 Vol. 06 No. 04, dengan judul **"Implementasi Data Mining K-Nearest Neighbor pada Penerimaan Karyawan Di PT Dwi Sumber Arca Waja"**. Rekrutmen merupakan langkah awal yang dilakukan oleh perusahaan untuk merekrut calon tenaga kerja profesional. Profesionalisme diperlukan sebagai salah satu input untuk menghasilkan output yang sesuai. Penelitian ini bertujuan untuk menerapkan metode klasifikasi *KNearest Neighbor (K-NN)* pada sistem seleksi rekrutmen. Metode K-NN akan menghitung tingkat kemiripan dengan mengukur jarak antara kebutuhan yang ditetapkan oleh perusahaan dengan data yang dimiliki oleh calon karyawan. Selanjutnya K-value diambil dari pelamar yang memiliki nilai *similarity* sebesar

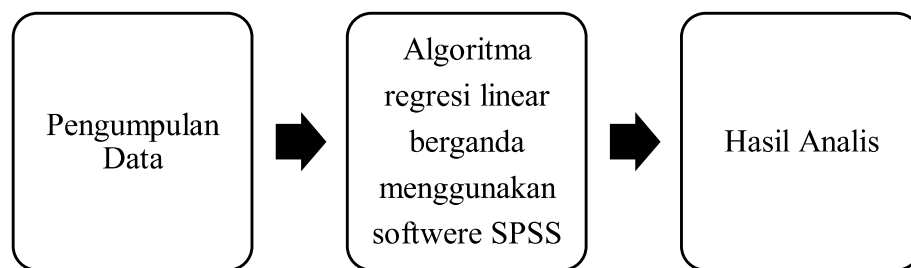
80%. Pelamar yang memiliki skor yang sama 80% akan diikutsertakan dalam tes berikutnya.

8. Penelitian oleh (Hutahaean & Handoko, 2022) dari Universitas Putera Batam, E-ISSN 2715-6265 Vol. 06 No. 05, dengan judul "**Penerapan Data Mining Untuk Memprediksi Penjualan Obat Di Klinik Harapan Kita Batam**". Penelitian ini bertujuan untuk mengklasifikasi obat yang dijual atau tidak dijual di klinik didasarkan pada variabel yang diperoleh dengan menggunakan *Naive Bayes Algorithm* yang mampu memberikan informasi tentang produk yang ada di klinik dan meminimalkan stok yang menumpuk untuk kategori tidak terjual. Ketersediaan data yang banyak dalam suatu wilayah kebutuhan informasi menjadi acuan dalam mengambil keputusan untuk membuat solusi bisnis. Informasi yang diinginkan dalam penelitian ini adalah untuk mendapatkan nilai akurasi data penjualan obat-obatan tersebut yang seringkali menjadi pilihan. Penelitian ini menggunakan tools Rapidminer sebagai media pengujian data yang akan diolah untuk mendapatkan nilai ROC (*Rank Order Centroid*) akurasi 80% berdasarkan kepentingan atau prioritas kriteria.

2.7 Kerangka Pemikiran

Kerangka Pemikiran merupakan suatu diagram yang menjelaskan rencana atau alur logika untuk berjalannya sebuah penelitian yang diangkat. Konsep di dalam kerangka pemikiran itu juga menjelaskan mengenai suatu hubungan antara

variabel yang satu dengan variabel yang lainnya. Melalui pemanfaatan data mining menggunakan metode regresi linier berganda di PT Mandiri Karya Nusantara dapat ditentukan susunannya, tahapan yang dibutuhkan untuk mendapatkan hasil rancangan dengan fungsi yang baik sebagai berikut:



Gambar 2. 4 Kerangka Pemikiran

Sumber: (Peneliti, 2022)

Seperti yang terlihat dari *flowchart* diatas, tahapan awal pada penelitian ini yaitu:

1. Dihimpun lalu dilaksanakan cleaning data, transformation guna mengetahui pola pengambilan data. Tahapan tersebut menghasilkan variabel-variabel masukan yang akan digunakan.
2. Tahap ini ialah proses pengkategorisasi data dengan memanfaatkan metode regresi linier berganda. Selanjutnya dilakukan pengujian dengan *software* SPSS (*Statistical Product and Service Solution*) berdasarkan dari data-data yang diperoleh.
3. Temuan penelitian ini adalah aturan untuk membuat keputusan dalam bentuk aturan yang digunakan sebagai panduan ketika menentukan kinerja kerja karyawan.