

## BAB 2

### TINJAUAN PUSTAKA

Bab ini membahas tentang beberapa teori yang diambil menurut pendapat dan pandangan para ahli tentang konsep proposal ini yang digunakan sebagai bahan acuan dan dasar pandangan dalam penelitian. Tinjauan Pustaka berisikan konsep-konsep mengenai permasalahan yang diteliti. Yang dimaksud konsep ialah unsur dasar terbentuknya teori sebagai landasan, dan menjadikan penelitian ini memiliki batasan. Batasan tersebut dibuat agar isi dari proposal tidak menyimpang dari kajian yang dibahas. Adapun komponen yang dibahas pada bab ini, yaitu: (1) *Data Mining*, (2) *Knowledge Discovery from Data (KDD)*, (3) *Clustering*, (4) *K-means*.

#### **2.1 Data Mining**

Berikut ini merupakan beberapa pengertian mengenai *Data Mining* menurut para ahli :

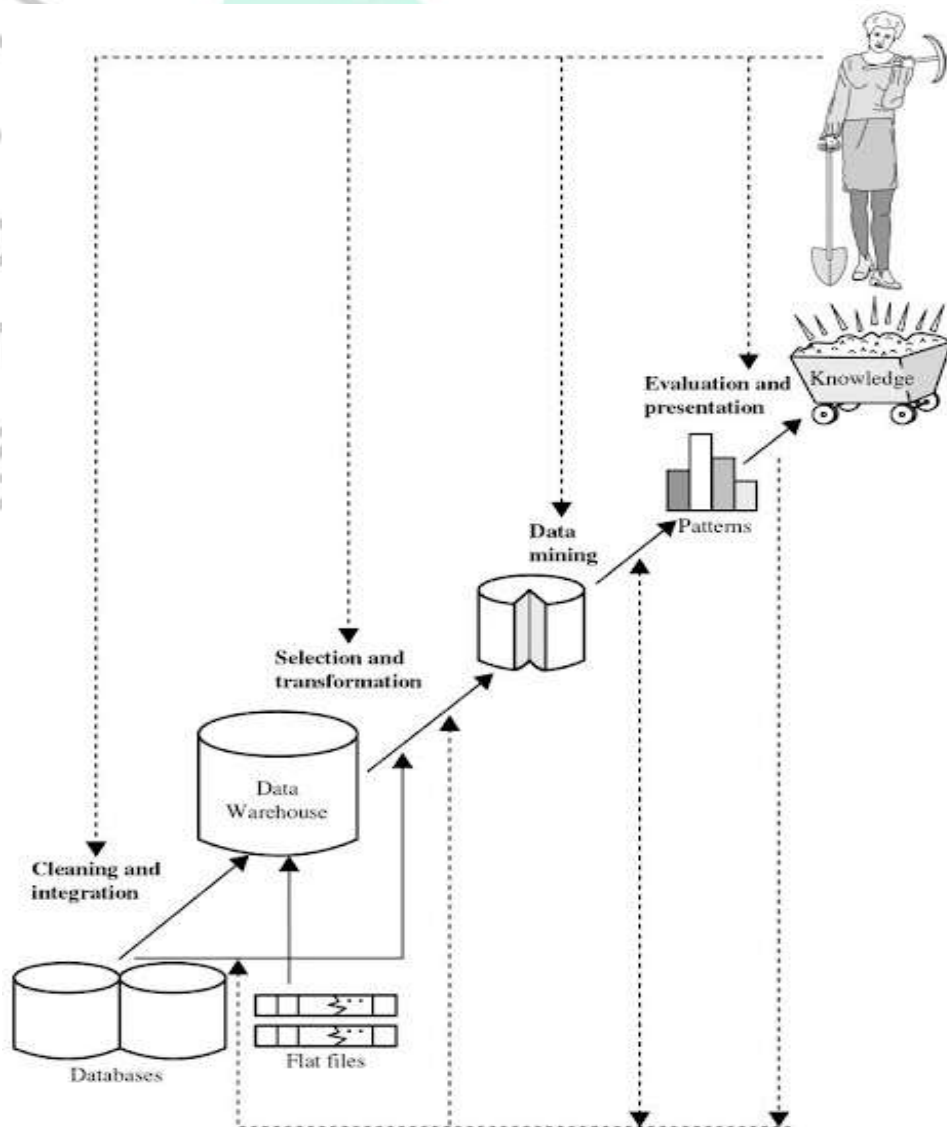
- Menurut (Suntoro, 2019), *Data Mining* merupakan sebuah proses untuk mendapatkan informasi yang berguna dari basis data yang besar dan perlu diekstraksi agar menjadi informasi baru yang dapat digunakan untuk membantu dalam pengambilan keputusan.
- Menurut (Werdiningsih, 2020:17), *Data Mining* merupakan sebuah bidang ilmu yang berguna untuk menangani masalah dalam pengambilan informasi dari *data warehouse* dengan menggabungkan beberapa teknik dari static, pembelajaran mesin, visualisasi data, pengenalan pola dan database.
- Menurut (Jaggia, 2020), *Data Mining* merupakan kompleks untuk memeriksa suatu kumpulan besar data dengan tujuan menemukan pola yang tersembunyi dan digunakan untuk mendapatkan pengetahuan tentang bisnis yang berharga

Dari keterangan diatas, dapat disimpulkan jika *data mining* adalah suatu

langkah penting dalam menemukan suatu pengetahuan pada proses *Knowledge Discovery from Data* (KDD) yang merupakan proses menemukan pengetahuan dalam database. Pada proses KDD ini dilakukan proses mengekstraksi atau identifikasi suatu pola, pengetahuan dan informasi yang potensial dari kumpulan besar data. Pengetahuan atau informasi yang dihasilkan dari proses KDD memiliki sifat sah, mudah mengerti, baru dan bermanfaat.

Tahapan proses *Knowledge Discovery from Data* (KDD) ini tentu memerlukan data. Data yang digunakan untuk KDD merupakan data yang telah terpisah dari data operasional.

Langkah-langkah dari proses *Knowledge Discovery from Data* (KDD) dapat dilihat pada gambar dibawah ini :



Gambar 2.1 : Proses *Knowledge Discovery from Data* (KDD)

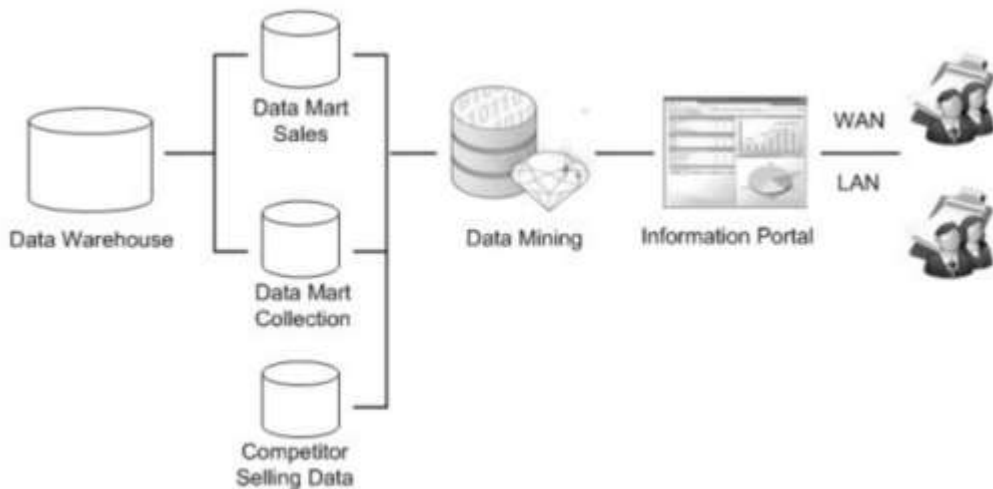
Sumber : Han, Jiawei *et al.* 2012

Berikut proses pada *Knowledge Discovery from Data* (KDD) secara bertahap menurut (Han, J., et al. 2012:6-8) :

- *Data Cleaning* (Pembersihan data)  
Pada tahap ini bertugas untuk menghapus atau menghilangkan *noise* dan data yang sering berubah atau tidak konsisten. Di tahap ini juga, data-data yang tidak memiliki kelengkapan yang dibutuhkan juga dihapus.
- *Data Integration* (Integrasi data)  
Proses dimana menggabungkan data-data berulang yang berasal dari berbagai sumber dan digabungkan menjadi satu.
- *Data Selection* (Seleksi data)  
Tahap ini bertugas untuk menyeleksi atau memilih data yang dianggap relevan dan berguna untuk analisis. Bertujuan untuk mengurangi atribut dan kompresi data.
- *Data Transformation* (Transformasi data)  
Proses dimana transformasi atau mengubah data terpilih ke dalam bentuk yang sesuai untuk *mining procedure* dengan meringkas atau menggabungkan operasi.
- *Data Mining* (Penambangan data)  
Proses ini merupakan proses inti dari proses KDD. Dalam proses ini data akan diolah dengan berbagai teknik dengan tujuan mencari informasi potensial yang nantinya akan menjadi data yang dapat digunakan. Pemilihan teknik dan algoritma sangat bergantung terhadap proses dan tujuan KDD keseluruhan.
- *Pattern Evaluation* (Evaluasi pola)  
Proses dimana mengidentifikasi pola-pola yang merupakan dasar pengetahuan pada tahap-tahap yang diberikan.
- *Knowledge Presentation* (Representasi pengetahuan)  
Tahap ini merupakan proses paling akhir dari keseluruhan proses KDD. Data-data atau penemuan pengetahuan kemudian direpresentasikan atau

divisualisasikan agar pengguna mudah memahami hasil dari *data mining*.

Secara umum, terdapat beberapa metode *data mining* salah satunya adalah *Clustering* yang akan digunakan sebagai konsep pada penelitian ini. Menurut (Tan, 2006), *Clustering* merupakan suatu proses untuk mengelompokkan data dalam beberapa *cluster* (kelompok) hingga data dalam 1 kelompok memiliki tingkat kesamaan yang maksimum dan data antara satu *cluster* dengan *cluster* lain memiliki kesamaan yang minimum. *Clustering* digunakan untuk mengumpulkan data yang dibagi menjadi beberapa kelompok berdasarkan kemiripan atribut atau karakteristik yang dimilikinya. Contoh yang akan digunakan untuk penelitian ini adalah *Customer Segmentation* yaitu dengan mengelompokkan data konsumen menjadi beberapa kelompok berdasarkan tingkat kemiripannya.



Dibawah ini adalah contoh dari arsitektur data mining :

Gambar 2.2 : Contoh arsitektur Data Mining

Sumber : Darudiato, S & Anindito. 2013

Pada gambar diatas *Data Mart Sales*, *Data Mart Collection* dan *Competitor Selling Data* adalah data set yang diambil dari *data warehouse*.

## 2.2 K-means

Dalam penelitian ini metode *clustering* yang digunakan adalah *K-Means*. *K-Means* merupakan metode *clustering* yang umum dan tergolong sederhana. *K-Means* dapat mengelompokkan sejumlah data besar dalam waktu yang singkat. *K-Means* juga dapat diartikan sebagai suatu metode untuk menganalisa data atau metode dari *data mining* yang melakukan proses permodelan *unsupervised learning* dan menggunakan metode yang mengelompokkan data dari berbagai partisi. Tujuan *K-Means* adalah untuk meminimalisir fungsi objektif yang telah diatur dalam proses *clustering* dengan meminimalisir variasi data yang ada dalam kelompok (*cluster*) dan memaksimalkan variasi data pada kelompok (*cluster*) lainnya.

*K-Means* adalah sebuah algoritma dasar dan diterapkan sebagai berikut :

- Menentukan jumlah kelompok (*cluster*).
- Mendistribusikan data *cluster* secara acak.
- Menghitung rata-rata data yang ada di *cluster*.
- Menghitung jarak data dan nilai centroid (*K-means clustering*).

## 2.3 Analisis Media Sosial

Analisis Media Sosial merupakan proses mengumpulkan dan mengolah data berharga dari berbagai saluran media sosial dalam kurun waktu tertentu. Hasil analisa tersebut dapat digunakan untuk membantukita mengetahui klasifikasi target pasar, kekuatan pesaing bahkan membantu dalam membuat strategi bisnis.

### 3 Metode Analisis Media Sosial :

- Analisis Deskriptif, umumnya berupa statistik yang didapatkan dari proses perhitungan secara matematis. Fungsinya agar mempermudah dalam mengetahui perkiraan dari penyebaran sampel data dan populasi. Biasanya digunakan untuk mengetahui rincian aktivitas pengguna di media sosial.
- Analisis Prediktif, analisis ini mengidentifikasi resiko atau peluang berdasarkan rekam data yang dimiliki. Fungsinya untuk menentukan

strategi berdasarkan data perilaku pengguna media sosial. Analisis ini memungkinkan kita untuk mengolah data menjadi sebuah laporan yang nantinya digunakan untuk menunjang strategi berikutnya. Contohnya kita bisa mengetahui segala bentuk opini pengguna mengenai produk apakah cenderung positif/negatif.

- Analisis Preskriptif, analisis ini umumnya cenderung menampilkan rekomendasi terhadap sebuah kasus di media sosial.

Contoh analisis ini :

- a) Konten seperti apa yang mampu meningkatkan interaksi pengguna di media sosial?
- b) Produk seperti apa yang cocok di daerah A?

Analisis Media Sosial penting dilakukan untuk mengetahui resiko dan peluang yang dimiliki perusahaan di media sosial. Sehingga, strategi yang akan dijalankan berikutnya akan lebih akurat dan sesuai dengan perhitungan yang tepat. Dan pada umumnya analisis diatas akan dilakukan menggunakan alat (*tools*) agar hasil lebih akurat dan mempermudah kita dalam melakukan proses pengumpulan dan pengolahan data di media sosial.

#### **2.4 Flowchart**






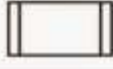


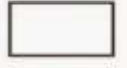




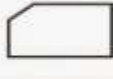


Menurut Nurmalina (2017 : 86), *Flowchart* merupakan sebuah representasi secara simbolik dari sebuah algoritma atau prosedur yang digunakan untuk menyelesaikan suatu permasalahan, *flowchart* dapat digunakan untuk melakukan pengecekan terhadap bagian yang terlewatkan dalam analisis data.

Fungsi *Flowchart* :

- Membantu menggambarkan alur proses dari sistem sehingga lebih mudah dipahami.
- Membantu dalam mengidentifikasi permasalahan yang mungkin terjadi pada suatu sistem, seperti alur proses yang tidak sesuai.
- Membantu meningkatkan efisiensi dan mengoptimalkan alur proses dari suatu sistem.
- Dapat digunakan sebagai panduan untuk mengembangkan sistem.

- Dapat mempermudah komunikasi antar anggota karena dapat menjelaskan alur proses dengan lebih mudah dipahami.

Simbol-simbol *Flowchart* :

	<b>Flow Direction symbol</b> Yaitu simbol yang digunakan untuk menghubungkan antara simbol yang satu dengan simbol yang lain. Simbol ini disebut juga connecting line.		<b>Simbol Manual Input</b> Simbol untuk pemasukan data secara manual on-line keyboard
	<b>Terminator Symbol</b> Yaitu simbol untuk permulaan (start) atau akhir (stop) dari suatu kegiatan		<b>Simbol Preparation</b> Simbol untuk mempersiapkan penyimpanan yang akan digunakan sebagai tempat pengolahan di dalam storage.
	<b>Connector Symbol</b> Yaitu simbol untuk keluar - masuk atau penyambungan proses dalam lembar / halaman yang sama.		<b>Simbol Predefine Proses</b> Simbol untuk pelaksanaan suatu bagian (sub-program)/prosedure
	<b>Connector Symbol</b> Yaitu simbol untuk keluar - masuk atau penyambungan proses pada lembar / halaman yang berbeda.		<b>Simbol Display</b> Simbol yang menyatakan peralatan output yang digunakan yaitu layar, plotter, printer dan sebagainya.
	<b>Processing Symbol</b> Simbol yang menunjukkan pengolahan yang dilakukan oleh komputer		<b>Simbol disk and On-line Storage</b> Simbol yang menyatakan input yang berasal dari disk atau disimpan ke disk.
	<b>Simbol Manual Operation</b> Simbol yang menunjukkan pengolahan yang tidak dilakukan oleh computer		<b>Simbol magnetik tape Unit</b> Simbol yang menyatakan input berasal dari pita magnetik atau output disimpan ke pita magnetik.
	<b>Simbol Decision</b> Simbol pemilihan proses berdasarkan kondisi yang ada.		<b>Simbol Punch Card</b> Simbol yang menyatakan bahwa input berasal dari kartu atau output ditulis ke kartu
	<b>Simbol Input-Output</b> Simbol yang menyatakan proses input dan output tanpa tergantung dengan jenis peralatannya		<b>Simbol Dokumen</b> Simbol yang menyatakan input berasal dari dokumen dalam bentuk kertas atau output dicetak ke kertas.

Gambar 2.3 : Simbol-simbol *Flowchart*

Sumber : *John W. Satzinger et al.* 2016

## 2.5 Shopee



Gambar 2.4 : Logo Shopee

Sumber : [Shopee.co.id](http://Shopee.co.id)

Shopee merupakan sebuah salah satu *e-commerce* ternama di Indonesia. Shopee didirikan pada tahun 2009 oleh Forrest Li dan memiliki kantor pusat di Singapura yang dimiliki oleh Sea Limited (dulunya adalah Garena). Shopee pertama kali diluncurkan pada tahun 2015 di Singapura yang kemudian meluas ke berbagai negara di Asia diantaranya Thailand, Taiwan, Vietnam, Filipina, Malaysia hingga sampai ke Indonesia.

Di Indonesia, Shopee Indonesia dipimpin oleh Chris Feng dimana Beliau telah memiliki banyak pengalaman dibidang *e-commerce* lainnya seperti di Zalora dan Lazada. Terbukti dengan dibawah kepemimpinan Beliau, Shopee meraih banyak prestasi diantaranya: pada tahun 2017, TheAsianParent mengungkapkan jika Shopee merupakan platform berbelanja pilihan pertama dari para ibu di Indonesia. Prestasi dan penghargaan lainnya seperti “*The Indonesian Netizen Brand Choice Award 2017*” pada Maret 2017 oleh Warta Ekonomi, “*The Best in Marketing Campaign*” pada September 2017 oleh Majalah Marketing, “*Bright Awards Indonesia 2017*” yang mengesal 6 nominator lainnya. Tidak heran Shopee menduduki peringkat *e-commerce* nomer 1 di Indonesia berdasarkan hasil survei



yang dilakukan di Snapcart dan App Annie dengan jumlah unduhan terbanyak di Google Playstore dan Apple Store.

## 2.6 Orange Data Mining



Gambar 2.5 : Logo *Orange Data Mining*

Sumber : [Orangedatamining.com](http://Orangedatamining.com)

Orange atau yang lebih dikenal dengan sebutan Orange Data Mining merupakan sebuah perangkat lunak open source atau software yang digunakan untuk melakukan proses data mining atau data analytic melalui konsep visual programming. Orange menampilkan front-end pemrograman visual untuk analisis data. Orange Data Mining dikembangkan oleh University of Ljubljana dengan tujuan untuk membuat software yang dapat mengolah data secara visual tanpa perlu pengalaman membuat program (dapat dilakukan tanpa membuat codingan). Orange Data Mining memungkinkan seseorang untuk melakukan data mining tanpa harus membuat kode program sehingga mempercepat dan mempermudah proses analisis.

## 2.7 Tinjauan Studi

Hal yang menjadi acuan disusunnya proposal ini karena saat ini perusahaan *e-commerce* berlomba-lomba untuk memberikan pelayanan yang terbaik dengan berbagai cara. Berikut merupakan beberapa penelitian yang

memiliki korelasi yang mengarah dengan penelitian yang dibahas pada proposal ini seperti :

- 1) Penelitian pertama yang dijadikan referensi berjudul “**Segmentasi Pelanggan Menggunakan Clustering K-Means**” yang disusun oleh Indri Yuli Kurniawati, 2018. Penelitian ini dibuat dikarenakan meningkatnya persaingan pasar dalam bidang perdagangan khususnya dalam usaha pengadaan kebutuhan interior, maka diperlukan pengelolaan data dan strategi bisnis untuk mempertahankan hubungan yang baik dengan pelanggan terutama pelanggan lama yang loyal. Metode yang digunakan Penulis adalah metode waterfall dengan proses Analisa, Design, Code, Testing, Penerapan dan Penelitian. Dalam penelitian ini membahas bahwa metode Clustering K-Means dapat digunakan untuk mengelola pesanan dan data pelanggan guna membantu perusahaan dalam membuat strategi pemasaran untuk mempertahankan pelanggan lama maupun baru. Data diambil dari transaksi selama 1 bulan dengan melibatkan total transaksi, total jumlah pembelian barang dan frekuensi transaksi dan diolah menggunakan metode Clustering K-Means hingga membentuk beberapa kelompok untuk mengetahui tingkat loyalitas pelanggan. Dengan hasil pengelompokan tersebut diharapkan perusahaan dapat menentukan strategi bisnis yaitu memberikan potongan harga bagi pelanggan yang memiliki loyalitas tinggi.
- 2) Penelitian kedua dengan judul “**Implementasi Algoritma K-Means Clustering Menggunakan Orange Untuk Penentuan Produk Busana Muslim Terlaris**” yang disusun oleh M Syukron Nawawi, Falentino Sembiring & Adhitia Erfina, 2021. Rancangan ini dibuat untuk membantu perusahaan dalam menentukan produk laris, guna untuk mengoptimalkan persediaan barang dan meningkatkan promosi penjualan. Rancangan ini menggunakan algoritma K-Means untuk melakukan pengelompokan atau clusterisasi pada data yang di uji. Penelitian ini menggunakan metode Clustering K-Means dan implementasi program dengan menggunakan

software Orange Data Mining untuk mempercepat dan mempermudah proses clustering untuk mengetahui mana produk yang paling laris. Hasil penelitian ini terdapat 3 kelompok kategori yaitu kelompok 1 yang berisikan produk kurang laris, kelompok 2 yang berisikan produk yang cukup laris dan terakhir kelompok 3 yang berisikan produk sangat laris. Dari hasil penelitian inilah yang nantinya akan membantu perusahaan membuat strategi untuk meningkatkan penghasilan dan mengoptimalkan promosi penjualan.

- 3) Penelitian berjudul **“Analisis Text Clustering Akun Fanpage Shopee Indonesia Dengan Komentar Followers Menggunakan Tools Orange Data Mining”** yang disusun oleh Ayuni Sentiya & Heri Suroyo, 2019. Rancangan ini dibuat untuk mengetahui sifat dan minat para followers dengan menyamakan data yang diambil melalui status dan komentar pada *fanpage* Shopee Indonesia yang ada di Twitter guna membuat strategi bisnis, dengan metode text mining untuk penggalian informasi. Penelitian ini menggunakan metode Clustering K-Means untuk mengelompokkan data yang diambil dari review dalam bentuk status dan komentar dari postingan akun twitter Shopee Indonesia dalam 3 bulan terakhir. Kemudian hasilnya akan dianalisis menjadi cluster hingga akhirnya menghasilkan kata yang sering (dominan) muncul dari status dan komentar. Kesimpulan dari hasil penelitian ini akan memberikan gambaran pesan yang dominan muncul pada status dan komentar dari followers pada akun twitter Shopee Indonesia dan menjadi topik pembicaraan pada setiap kelompok (Cluster) yang dibuat. Hal ini dapat dimanfaatkan oleh perusahaan untuk lebih mengetahui minat dan pendapat dari followers-nya sehingga dapat lebih mengenal dan membuat strategi yang sesuai untuk menarik perhatian followers guna meningkatkan promosi Shopee Indonesia.
- 4) Penelitian berjudul **“Penerapan Data Mining Menggunakan Algoritma Apriori Terhadap Data Transaksi Penjualan Bisnis Ritel”** yang

disusun oleh Robby Takdirillah, 2020. Rancangan ini dibuat karena banyaknya permasalahan mengenai stok yang menumpuk hingga merugikan perusahaan maka diperlukannya sebuah strategi penjualan yang tepat untuk lebih mengoptimalkan penjualan dan persediaan stok barang. Supaya data transaksi penjualan menjadi lebih efisien maka perlu diolah dengan menggunakan teknik *data mining*. Metode yang digunakan oleh penulis dalam merancang program tersebut adalah melalui pendekatan *knowledge discovery in database*, yang meliputi analisa data hingga penentuan algoritma apriori. Tujuan dari dilakukannya penelitian ini adalah untuk mendapatkan informasi mengenai keterikatan antar produk yang sekiranya mendukung strategi penjuan dari data transaksi penjualan. Informasi tersebut nantinya akan digunakan untuk mengambil keputusan strategi penjualan. Hasil dari algoritma apriori juga dapat digunakan untuk mengolah data dari transaksi penjualan menjadi sebuah informasi baru dalam konteks keterikatan antar produk yang diambil dari pengujian dengan menggunakan *tools* Orange. (Robby Takdirillah, 2020).

- 5) Penelitian berjudul “**Implementasi *Unsupervised Learning* Pada Nilai Jasmani Kesamaptaan Sekolah Polisi Negara Dengan Metode *Clustering Analysis***” yang disusun oleh Zaenal Muttaqin, Donny Fernando & Selvia Sulastriani, 2023. Rancangan ini dibuat untuk menentukan pembentukan personil kepolisian dimana nantinya akan dikelompokkan berdasarkan hasil nilai ujian fisik yang telah dilakukan. Jika siswa yang memiliki nilai tinggi dalam ujian fisik maka akan mendapatkan binaan khusus dan kemungkinan besar akan masuk ke Satuan Brimob, sedangkan yang memiliki nilai ujian lebih kecil akan dimasukkan ke kesatuan Polda ataupun Polres. Selama ini pengolahan nilai masih dilakukan secara manual sehingga mengakibatkan terjadinya penilaian secara subjektif oleh pihak pimpinan. Maka dari ini diperlukan sebuah pengelompokan nilai berdasarkan data dengan menerapkan *Unsupervised Learning* dengan metode *Hierarchical Clustering*. Data akan di observasi dan dilakukan analisis data hasil ujian dengan

menggunakan metode *Unsupervised Learning Clustering Analysis*. Setelah dianalisis maka grafik yang dihasilkan akan ditampilkan dalam bentuk *Visualisasi Data*. Dari hasil *Visualisasi Data* inilah yang nantinya akan digunakan petugas untuk menentukan dan mengambil keputusan kepada siswa kesamaptaan.

