

BAB II

LANDASAN TEORI

2.1 PRISMA

PRISMA (Preferred Reporting Items for Systematic Reviews And Meta-Analyses) merupakan alat dan panduan yang digunakan untuk melakukan penelitian terhadap sebuah systematic reviews dan meta analysis (Mohamed et al., 2020). Menurut Mohamed et al., (2020), Prisma diagram adalah Diagram alir menggambarkan aliran informasi melalui fase yang berbeda dari tinjauan sistematis. Ini memetakan jumlah catatan yang diidentifikasi, disertakan dan dikecualikan, dan alasan pengecualian. Templat yang berbeda tersedia tergantung pada jenis ulasan (baru atau diperbarui) dan sumber yang digunakan untuk mengidentifikasi studi . Pembuatan Systematic Literature Review (PRISMA) terdiri dari 4 langkah, yaitu :

1. *Identification* (Identifikasi) journal yang akan disertakan dalam meta-analisis.
2. *Screening* (Penyaringan), penyaringan atau pemilihan data.
3. *Eligibility* (Kelayakan), menentukan artikel yang akan dijadikan bahan literature review
4. *Included* (Inklusi), yakni penggabungan dan pelaporan hasil.

2.1.1 Identification

Identification (Identifikasi) adalah kegiatan yang bertujuan untuk memeriksa dan menganalisa, dalam proses ini penulis mengidentifikasi masalah penelitian (*research problem*) dikarenakan hal tersebut menentukan kualitas suatu penelitian (Mohamed et al., 2020).

2.1.2 Screening

Screening (Penyaringan) adalah sebuah proses penyaringan untuk memasukkan atau mengecualikan artikel berdasarkan kriteria yang ditentukan oleh penulis (Mohamed et al., 2020).

2.1.3 Eligibility

Eligibility adalah tahap penentuan kelayakan artikel semua artikel diperiksa dengan membaca judul, abstrak, hasil dan pembahasan untuk memastikan memenuhi kriteria inklusi dan sejalan dengan tujuan penelitian saat ini (Mohamed et al., 2020).

2.1.4 Included

Included merupakan tahap terakhir dari PRISMA (*Preferred Reporting Items for Systematic Reviews And Meta-Analyses*) flow diagram, pada tahap ini kriteria inklusi adalah dimana artikel keluar dan memenuhi persyaratan untuk di analisis (Mohamed et al., 2020).

2.2 Tahapan Systematic Reviews

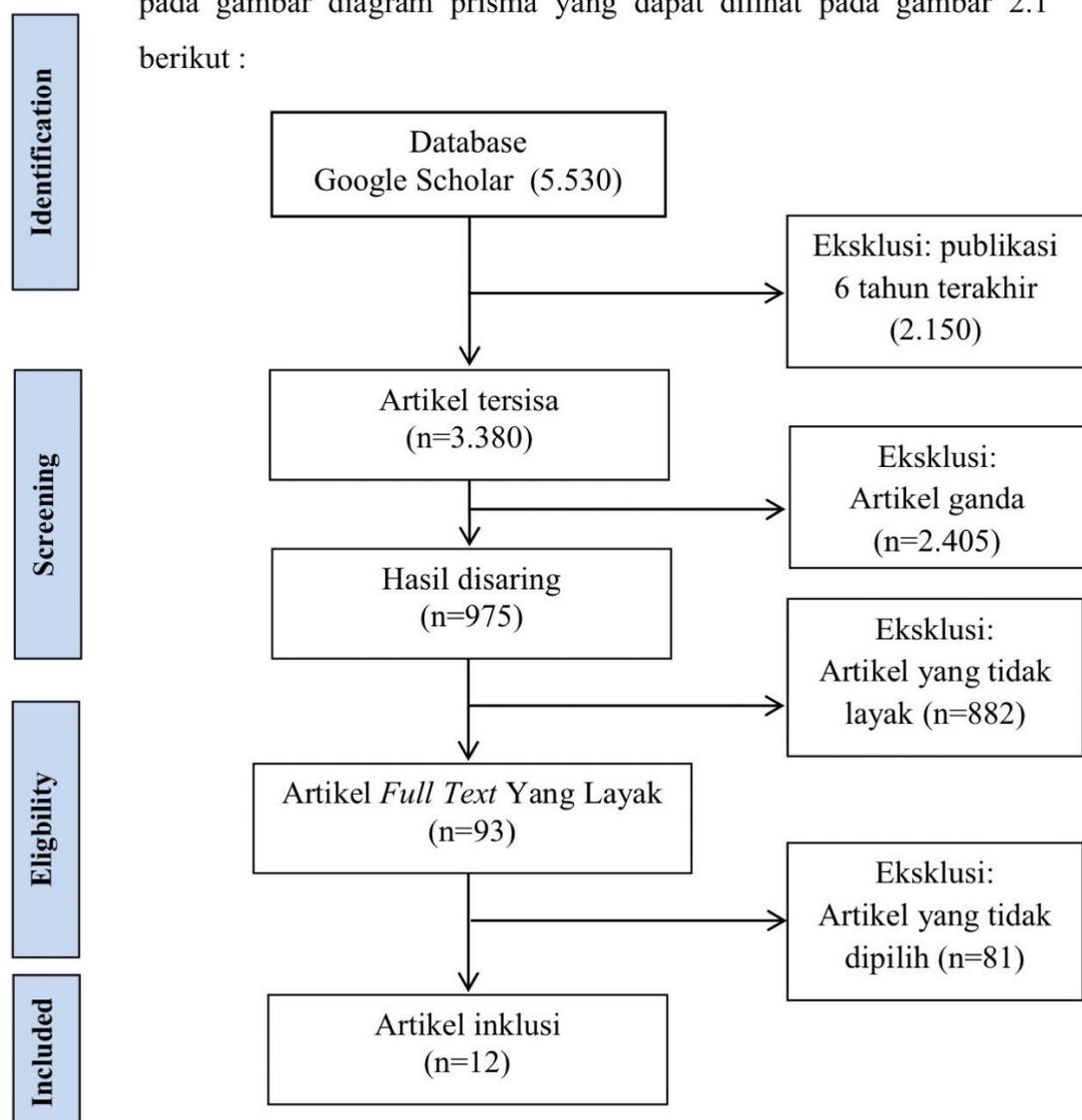
Menurut Oxman Systematic Reviews adalah metodologi penelitian yang mewakili investigasi topik tertentu, menekankan satu pertanyaan yang telah diidentifikasi, dievaluasi, dipilih, dan diputuskan secara sistematis sesuai dengan kriteria yang telah ditetapkan sebelumnya, berdasarkan bukti penelitian yang berkualitas tinggi yang terkait dengan pertanyaan penelitian (Adisasmito, 2007).

Dalam melakukan tahapan *systematic review* pada penelitian ini penulis mengumpulkan data yang terdapat pada database google scholar dan melakukan beberapa tahapan yaitu:

1. Identification, dalam tahap ini penulis mengumpulkan artikel sebanyak 5.530.
2. Screening, kemudian dalam tahap ini penulis memfilter artikel dari tahun 2016-2022 sehingga terkumpul artikel sebanyak 3.380 yang memiliki fokus kajian pada penerapan data *mining* dalam pengelompokan penyebaran data mahasiswa baru untuk strategi promosi, namun ketika di screening kembali ternyata terdapat 2.405 artikel yang terdeteksi ganda sehingga peneliti menghapusnya menjadi 975.
3. Eligibility, dalam tahap ini penulis memproses kelayakan yang disaring kembali dari 975 artikel yang sesuai dengan riset, *fulltext* dan

artikel memiliki bahasan mengenai penerapan data *mining* dalam pengelompokan penyebaran data mahasiswa baru untuk strategi promosi ditemukan sebanyak 93 artikel yang sesuai .

4. Included, ini merupakan tahap terakhir dari tahapan systematic reviews Temuan pada artikel yang memenuhi kelayakan yaitu sebanyak 93 artikel tetapi terdapat eksklusi, sehingga penulis memilih 12 artikel saja yang akan menjadi referensi dalam penelitian ini sesuai pada gambar diagram prisma yang dapat dilihat pada gambar 2.1 berikut :



Gambar 2.1 Diagram PRISMA : Tahapan Systematic Review

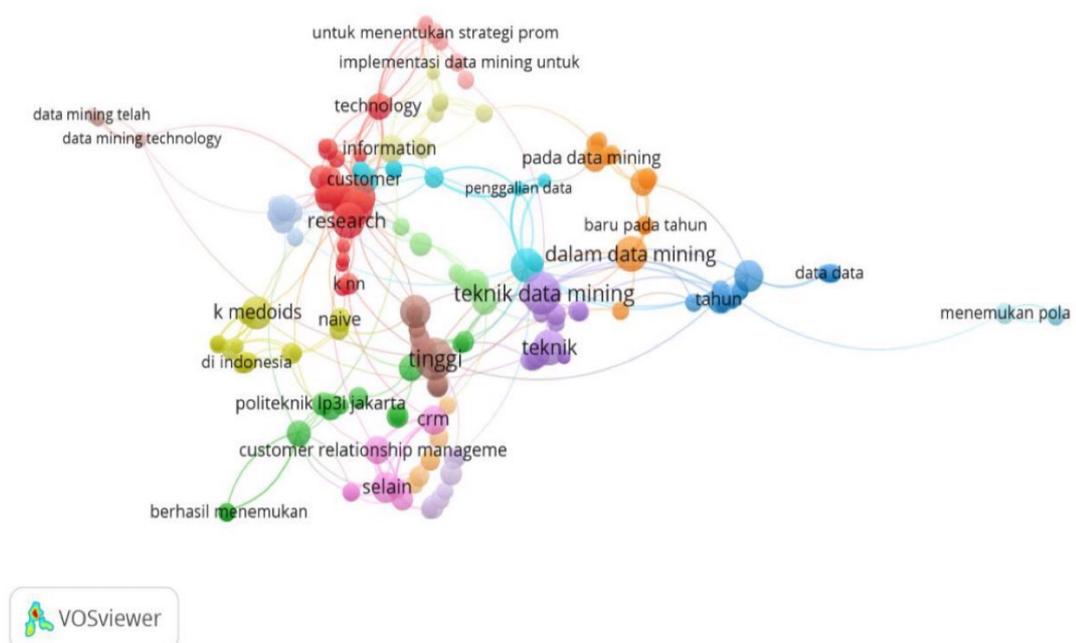
2.3 VOS Viewer

VOS viewer merupakan perangkat lunak untuk pemetaan berbagai macam jenis analisis bibliometrik. Menurut Pattah dalam (Sarman & Soebagyo, 2022) Bibliometrik yaitu metode yang dipakai dalam sebuah artikel yang bersifat deskriptif untuk mengetahui pengarangnya, tingkat kolaborasi serta literturnya. Sedangkan menurut Winardi et al., Bibliometrik adalah ilmu yang membahas penulisan serta analisis yang matematis, Maka dapat disimpulkan bibliometrik adalah pustaka dalam publikasi ilmiah untuk mengetahui analisis penulisan, pengarang dan literatur (Sarman & Soebagyo, 2022).

Dari tahapan systematic review dapat dilihat pada gambar 2.1 penulis mengumpulkan hasil artikel sebanyak 5.030 yang bersumber dari database google scholar namun pengumpulan hasil artikel menjadi 975 ketika melalui tahap skrining, 975 artikel tersebut dapat divisualisasikan pemetaannya dengan VOS Viewer sebagai berikut :

a. Network Visualization

Opsi network *Visualization* pada VOS viewer memperlihatkan jejaring antar term saling berhubungan yang divisualkan sebagai berikut :

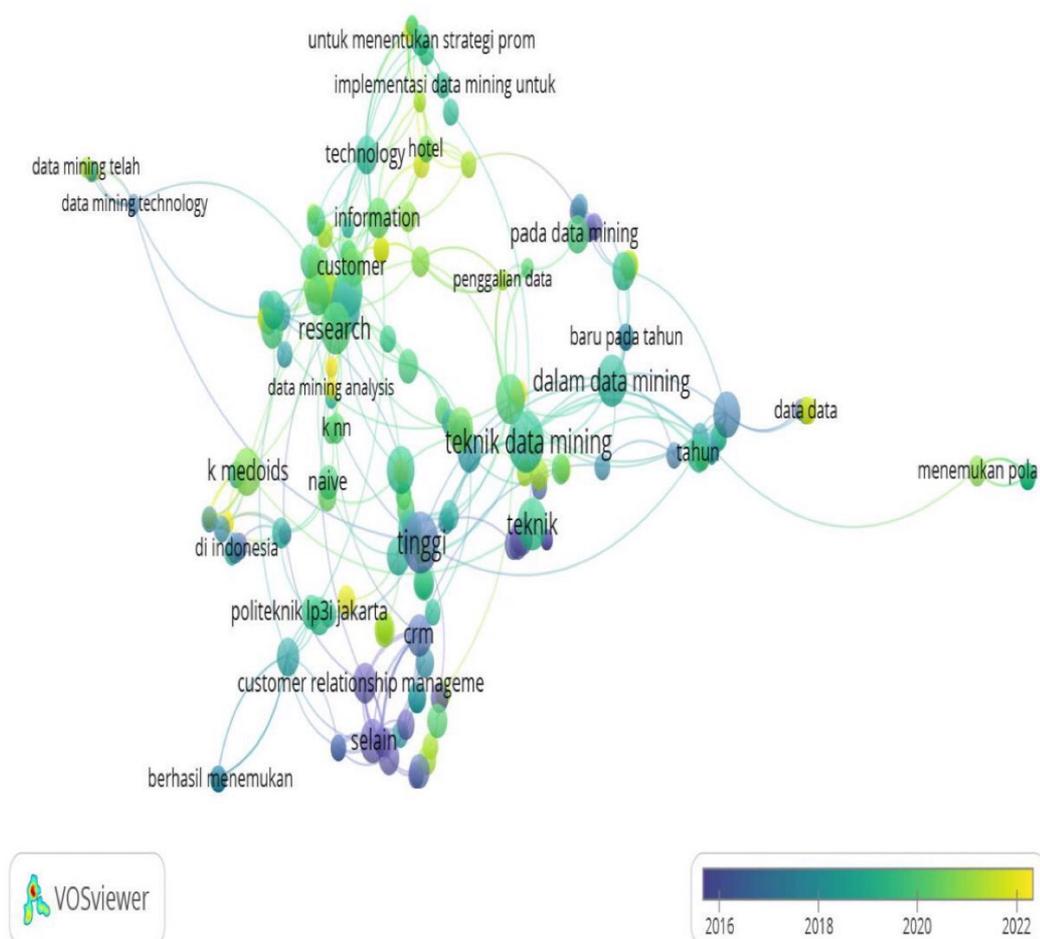


Gambar 2.2 Network Visualization Vos Viewer

Gambar 2.2 menunjukkan bahwa berdasarkan kata kunci (co-word), pemetaan pada kata kunci penerapan data mining dalam pengelompokan penyebaran data mahasiswa baru untuk strategi promosi membentuk menjadi 17 kluster 187 items terms yang saling berhubungan.

b. Overlay *Visualization*

Opsi overlay *Visualization* pada VOS Viewer memperlihatkan jejak history penelitian sehingga dapat dilihat berikut :

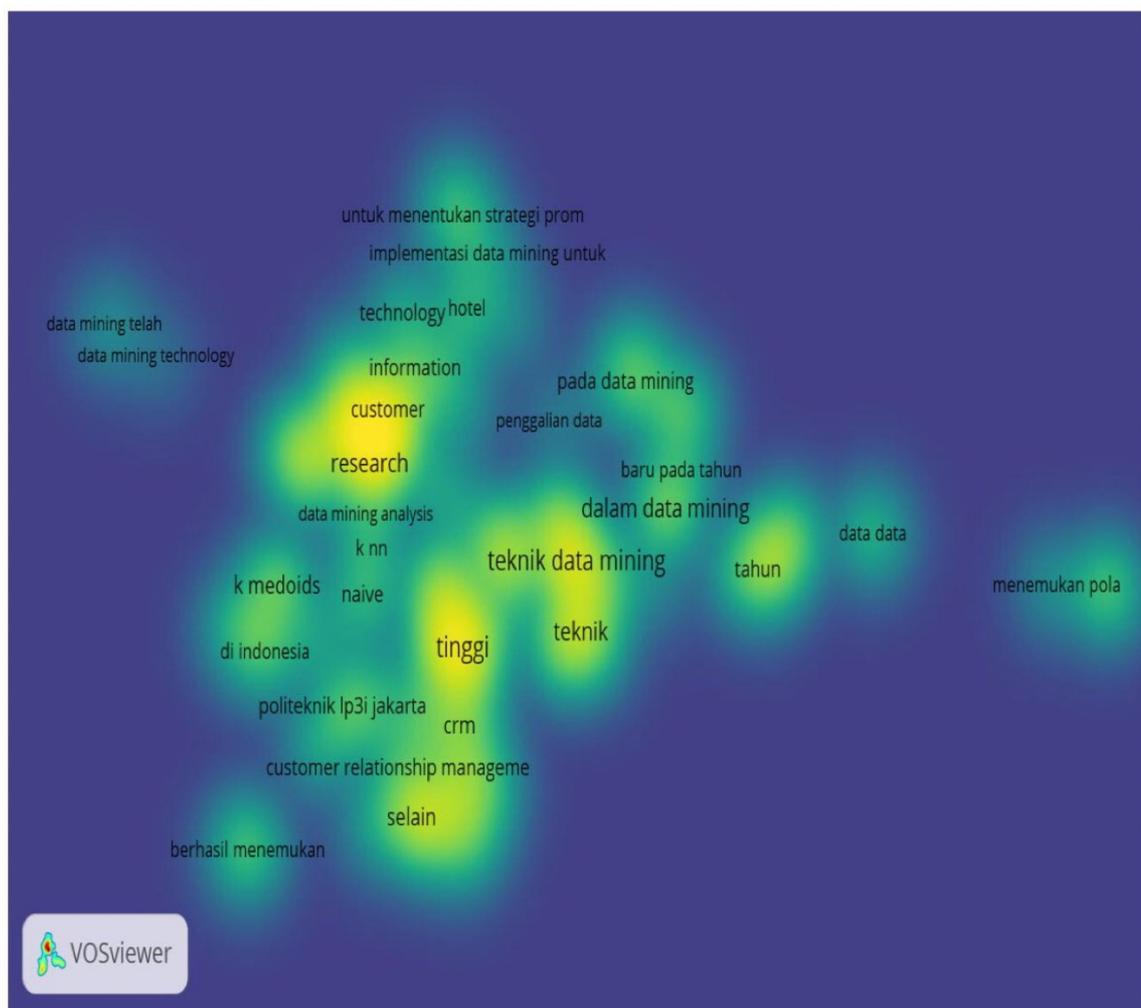


Gambar 2.3 Overlay *Visualization* Vos Viewer

Gambar 2.3 menunjukkan bahwa berdasarkan kata kunci (co-word), pemetaan pada kata kunci penerapan data mining dalam pengelompokan penyebaran data mahasiswa baru untuk strategi promosi membentuk menjadi 17 kluster dan dapat dilihat jejak historynya dari tahun ke tahun.

c. Density Visualization

Opsi *density Visualization* pada VOS viewer memperlihatkan penekanan pada kelompok penelitian



Gambar 2.4 Density Visualization Vos Viewer

Gambar 2.4 menunjukkan bahwa berdasarkan kata kunci (co-word), pemetaan pada kata kunci penerapan data mining dalam pengelompokan penyebaran data mahasiswa baru untuk strategi promosi membentuk

menjadi 17 kluster dan dapat dilihat kata kuncinya semakin terlihat terang maka kata kunci tersebut sudah banyak yang meneliti.

Dapat disimpulkan bahwa hasil pemetaan bibliometrik penggunaan data *mining* dengan pencarian kata kunci “Penerapan Data Mining Dalam Pengelompokan Penyebaran Data Untuk Strategi Promosi” didalam database google scholar setelah melalui proses screening dari tahun 2016-2022 menjadi 3.380 dan mengeksklusi artikel ganda menjadi 975 artikel mendapat hasil 17 kluster 187 items terms yang berhubungan.

2.4 Literature Review

Literature review adalah karya akademis yang berfokus pada topik tertentu (Cahyono et al., 2019). Menurut Rowley literature review memungkinkan penulis untuk melakukan identifikasi teori dengan metode mengembangkan teori atau metode Identifikasi kesenjangan yang ada di antara teori-teori yang relevan dengan hasil lapangan/penelitian (Cahyono et al., 2019). Menurut Cronin et al., dalam (Cahyono et al., 2019) Isi dari literature review terkadang bervariasi dari literature review ke literature review lain, tetapi proses untuk melakukan literature review akan mengikuti langkah dan tahapan tahapan yang serupa.

Dibawah ini merupakan tabel hasil literature review sebagaimana yang bisa dilihat pada tabel 2.1 berikut :

Tabel 2.1 Literature Review

No	Judul Artikel	Judul Jurnal	Nama Penulis, Tahun	Sumber	Metode	Tujuan Penelitian	Open Source Dataset	Hasil
1	PENERAPAN METODE <i>K-MEANS CLUSTERING</i> DALAM MENENTUKAN STRATEGI PROMOSI BERDASARKAN DATA PENERIMAAN MAHASISWA BARU	Jurnal TEKNO KOMPAK	Nanda Ayu Rahmalinda, Arief Jananto (2022)	Google Scholar	Algoritma <i>K-means Clustering</i>	Untuk menghasilkan pengelompokan data mahasiswa baru menggunakan teknik data mining berdasarkan profil mahasiswa baru yang mendaftar di Sekolah Tinggi Ilmu Ekonomi Assholeh Pernalang.	Dataset Penerimaan Mahasiswa Baru dengan Tahun Ajaran 2018/2019, 2019/2020, 2020/2021, dan 2021/2022	Penelitian ini menghasilkan tiga klasterisasi yaitu Cluster 1 berjumlah 189 data dengan presentase 42%, Cluster 2 berjumlah 186 data dengan presentase 41% dan Cluster 3 berjumlah 78 data dengan presentase 17%.

Tabel 2.1 Literature Review (Lanjutan)

No	Judul Artikel	Judul Jurnal	Nama Penulis, Tahun	Sumber	Metode	Tujuan Penelitian	Open Source Dataset	Hasil
2	IMPLEMENTASI ALGORITMA K-MEDOIDS UNTUK PENGELOMPOKAN SEBARAN MAHASISWA BARU	Jurnal Riset Sistem Informasi Dan Teknik Informatika (JURASIK)	Eka Irawan, Sandy Putra Siregar, Irfan Sudahri Damanik, Ilham Syaputra Saragih (2020)	Google Scholar	Algoritma K-Medoids	Agar dapat mengetahui dimana asal sekolah terbanyak dan tersedikit berdasarkan cluster tinggi dan rendah untuk melakukan promosi STIKOM Tunas Bangsa Pematangsiantar	Dataset BAK STIKOM Tunas Bangsa	Penelitian ini menghasilkan 3 cluster yakni klaster tinggi (0), klaster sedang (2), dan klaster rendah (1)

Tabel 2.1 Literature Review (Lanjutan)

No	Judul Artikel	Judul Jurnal	Nama Penulis, Tahun	Sumber	Metode	Tujuan Penelitian	Open Source Dataset	Hasil
3	PENERAPAN DATA MINING PADA PENERIMAAN MAHASISWA BARU DENGAN ALGORITMA <i>K-MEANS CLUSTERING</i>	Jurnal Teknik Informasi dan Komputer (TEKINKOM)	Septian Isnanto , Suryarini Widodo (2021)	Google Scholar	Algoritma <i>K-means Clustering</i>	Melakukan pengelompokkan data untuk mencari jurusan yang potensial	Dataset Politeknik STMI Jakarta tahun 2017 - 2020 sebanyak 1104	Penelitian ini menghasilkan beberapa cluster yang terdiri : Prodi administrasi bisnis otomotif menghasilkan 3 cluster, Prodi sistem informasi bisnis otomotif menghasilkan 2 cluster, Prodi teknik industri otomotif 2 cluster

Tabel 2.1 Literature Review (Lanjutan)

No	Judul Artikel	Judul Jurnal	Nama Penulis, Tahun	Sumber	Metode	Tujuan Penelitian	Open Source Dataset	Hasil
4	PENERAPAN DATA MINING MENGGUNAKAN ALGORITMA <i>K-MEANS CLUSTERING</i> UNTUK MENENTUKAN STRATEGI PROMOSI MAHASISWA BARU (Studi Kasus : Politeknik LP3I Jakarta)	JURNAL LENTERA ICT	Rony Setiawan (2017)	Google Scholar	Metode CRISP-DM dengan Algoritma <i>K-means Clustering</i>	Melakukan Pengelompokan Data agar bagian marketing dapat melakukan pemasaran dengan strategi yang tepat untuk mendapatkan calon mahasiswa	Dataset Politeknik LP3I Jakarta calon mahasiswa 5 tahun terakhir yaitu dari tahun 2010 sampai tahun 2015	Penelitian ini Menghasilkan jumlah kluster 4 ($k=4$) dengan cluster pertama 17007, cluster kedua 83, cluster ketiga sebanyak 12919 dan cluster keempat sebanyak 356 calon mahasiswa.

Tabel 2.1 Literature Review (Lanjutan)

No	Judul Artikel	Judul Jurnal	Nama Penulis, Tahun	Sumber	Metode	Tujuan Penelitian	Open Source Dataset	Hasil
5	PENERAPAN DATA MINING MENGGUNAKAN ALGORITMA KMEANS CLUSTERING PADA PENERIMAAN MAHASISWA BARU SEBAGAI METODE PROMOSI	JUTISAL (Jurnal Teknik Informatika Komputer Universal)	Nurafni Damanik, Mula Sigirot (2021)	Google Scholar	Algoritma <i>K-means Clustering</i>	Tujuan penelitian ini adalah salah satu dasar pengambilan keputusan untuk menentukan strategi dalam mempromosikan kampus	Dataset Institut Teknologi dan Bisnis Indonesia Kampus Milenial Medan dan Serdang Bedagai	Penelitian ini menghasilkan 3 cluster, yaitu 41 items cluster pertama, 52 items cluster kedua, dan 57 items cluster ketiga

Tabel 2.1 Literature Review (Lanjutan)

No	Judul Artikel	Judul Jurnal	Nama Penulis, Tahun	Sumber	Metode	Tujuan Penelitian	Open Source Dataset	Hasil
6	PENENTUAN STRATEGI PROMOSI PENERIMAAN MAHASISWA BARU DENGAN ALGORITMA CLUSTERING K-MEANS	JURNAL IC-Tech	Tria Titiani Chasanah, Widiyono (2017)	Google Scholar	Algoritma <i>K-means Clustering</i>	Tujuan dari penelitian ini adalah dapat menentukan informasi tujuan promosi berdasarkan peminatan program studi calon mahasiswa baru pada PMB STMIK Widya Pratama	Dataset STMIK Widya Pratama tahun 2014 dan 2015	Penelitian ini menghasilkan 3 kluster yang dapat dianalisa sebagai informasi untuk menentukan strategi promosi PMB STMIK Widya Pratama

Tabel 2.1 Literature Review (Lanjutan)

No	Judul Artikel	Judul Jurnal	Nama Penulis, Tahun	Sumber	Metode	Tujuan Penelitian	Open Source Dataset	Hasil
7	ANALISA SEBARAN MAHASISWA UNIVERSITAS STIKUBANK MENGGUNAKAN METODE K-MEANS	Jurnal Proceeding SENDIU	Felix Andreas Sutanto, Heribertus Yulianton, Jeffri Alfa Razaq (2020)	Google Scholar	Algoritma <i>K-means Clusterin</i> g dan Algoritma Lloyd	Tujuan dari penelitian ini adalah untuk melakukan pengelompokan terhadap data mahasiswa agar dapat diketahui kelompok-kelompok wilayah promosi yang dapat dijadikan prioritas untuk promosi	Dataset mahasiswa Unisbank pada tahun 2018 dan 2019	Penelitian ini menghasilkan 4 Cluster, Cluster 1 (48%), Cluster 2 (16%), Cluster 3 (14%), Cluster 4 (12%)

Tabel 2.1 Literature Review (Lanjutan)

No	Judul Artikel	Judul Jurnal	Nama Penulis, Tahun	Sumber	Metode	Tujuan Penelitian	Open Source Dataset	Hasil
8	PENERAPAN DATA MINING UNTUK MENENTUKAN LOKASI PROMOSI PENERIMAAN MAHASISWA BARU PADA UNIVERSITAS BANTEN JAYA (METODE <i>K-MEANS CLUSTERING</i>)	Jurnal ProTekInfo (Pengembangan Riset dan Observasi Teknik Informatika)	Ramdani Budiman, Rudianto (2019)	Google Scholar	Algoritma <i>K-means Clustering</i>	Tujuan dari penelitian ini adalah untuk menentukan lokasi atau daerah mana saja yang berpotensi untuk mendatangkan mahasiswa baru di perguruan tinggi	Dataset Mahasiswa Universitas Banten Jaya Tahun 2014	Penelitian ini terbentuk menjadi tiga cluster. Cluster pertama kategori peminat tinggi, cluster kedua kategori peminat rendah dan cluster tiga kategori peminat sedang

Tabel 2.1 Literature Review (Lanjutan)

No	Judul Artikel	Judul Jurnal	Nama Penulis, Tahun	Sumber	Metode	Tujuan Penelitian	Open Source Dataset	Hasil
9	PENYEBARAN MAHASISWA BARU MENGGUNAKAN METODE FUZZY C-MEANS UNTUK Mencari DAERAH PROMOSI YANG POTENSIAL	Journal of Information Technology and Computer Science (INTECOMS)	Mohd Rinaldi Amatha (2020)	Google Scholar	Algoritma Fuzzy C-Means	Tujuan dari penelitian ini adalah untuk mencari daerah potensial untuk strategi promosi dan juga dapat digunakan untuk mencari daerah binaan bagi STIKes Hang Tuah Pekanbaru	Dataset Data Mahasiswa STIKes Hang Tuah Pekanbaru Tahun 2016-2017	Penelitian ini terbentuk menjadi 3 cluster. Cluster pertama rendah (0,8676, 0,8232, 0,5427), cluster kedua sedang (0,9989, 0,8819, 0,6993, 0,7810) dan cluster tiga kategori peminat tinggi (0,7743, 0,6783)

Tabel 2.1 Literature Review (Lanjutan)

No	Judul Artikel	Judul Jurnal	Nama Penulis, Tahun	Sumber	Metode	Tujuan Penelitian	Open Source Dataset	Hasil
10	PENERAPAN DATA MINING MENGGUNAKAN ALGORITMA K-MEANS CLUSTERING PADA PENERIMAAN MAHASISWA BARU (STUDI KASUS : UNIVERSITAS ISLAM INDRAGIRI)	Jurnal SISTEMASI	Fitri Yunita (2018)	Google Scholar	Algoritma <i>K-means Clustering</i>	Tujuan dari penelitian ini adalah sebagai salah satu dasar pengambilan keputusan untuk menentukan strategi mempromosikan masing-masing program studi yang ada di universitas islam indragiri.	Dataset Universitas Islam Indragiri data mahasiswa baru tahun ajaran 2016-2017	Penelitian ini menghasilkan Cluster mahasiswa, yang terbentuk adalah tiga cluster, dengan cluster pertama 195 items, cluster kedua 271 items dan cluster ketiga sejumlah 50 items

Tabel 2.1 Literature Review (Lanjutan)

No	Judul Artikel	Judul Jurnal	Nama Penulis, Tahun	Sumber	Metode	Tujuan Penelitian	Open Source Dataset	Hasil
11	SEGMENTASI LOKASI PROMOSI PENERIMAAN MAHASISWA BARU MENGGUNAKAN METODE RFM DAN <i>K-MEANS CLUSTERING</i>	Jurnal Manajemen, Teknik Informatika, dan Rekayasa Komputer	Dyah Susilowati , Hairani , Hairani , Indah Puji Lestari , Khairan Marzuki , Lalu Zazuli Azhar Mardedi (2022)	Google Scholar	Algoritma <i>K-means Clustering</i>	Tujuan dari penelitian ini adalah melakukan segmentasi tingkat potensial sekolah sebagai strategi untuk memilih lokasi promosi penerimaan mahasiswa baru Universitas Bumigora.	Dataset data penerimaan mahasiswa baru Universitas Bumigora tahun 2019 dan 2020	Hasil penelitian ini adalah terbentuk 3 klaster Klaster sangat potensial (C2) terdapat 28 sekolah, klaster potensial (C3) terdapat 90 sekolah, dan klaster kurang potensial (C1) terdapat 152 sekolah.

Tabel 2.1 Literature Review (Lanjutan)

No	Judul Artikel	Judul Jurnal	Nama Penulis, Tahun	Sumber	Metode	Tujuan Penelitian	Open Source Dataset	Hasil
12	ANALISIS KELAYAKAN LOKASI PROMOSI DALAM PENERIMAAN MAHASISWA BARU (PMB) DENGAN ALGORITMA NAÏVE BAYES & DECISSION TREE C4.5	Jurnal Informatika	Sekar Wulandari, Mohammad Iqbal (2021)	Google Scholar	Algoritma Naïve Bayes dan Algoritma Decission Tree C4.5	Tujuan dari penelitian ini adalah untuk membantu institusi/lembaga dalam memperoleh rekomendasi kelayakan lokasi promosi berdasarkan dari beberapa pengukuran kriteria	Dataset penerimaan mahasiswa baru STMIK Bani Saleh Tahun angkatan 2015-2019 dan 2020-2021	Penelitian ini menghasilkan nilai akurasi pada algoritma Naïve Bayes sebesar 91,43% dan Decission Tree C4.5 sebesar 94,29%

Berdasarkan tabel 2.1 diatas, Terdapat 12 penelitian yang meneliti tentang beberapa algoritma untuk melakukan pengelompokan diantaranya ada algoritma *k-means*, *k-medoids* dan *fuzzy c-means*. Beberapa penelitian lain yang meneliti tentang bagaimana membuat strategi promosi untuk penerimaan mahasiswa baru agar promosi yang dibuat tepat sasaran ada berbagai macam algoritma yang digunakan, diantaranya ada algoritma *k-means*, *naive bayes*, *c4.5*, dan *lloyd*.

Pada penelitian Fitri Yunita yang berjudul PENERAPAN DATA MINING MENGGUNAKAN ALGORITMA K-MEANS CLUSTERING PADA PENERIMAAN MAHASISWA BARU (STUDI KASUS : UNIVERSITAS ISLAM INDRAGIRI). Penelitian ini melakukan pengelompokan terhadap data penerimaan mahasiswa baru di Universitas Islam Indragiri dengan memanfaatkan proses data mining dengan menggunakan teknik Clustering penelitian ini bertujuan untuk mencari lokasi strategi promosi yang tepat berdasarkan parameter Asal sekolah, Program studi, dan nilai UAN dari 516 data mahasiswa. dari hasil penelitian Fitri Yunita ini penulis membentuk 3 cluster, cluster pertama 195 items, cluster kedua 271 items, dan cluster ketiga 50 items.

Adapun tujuan penelitian ini bahwa dalam melakukan pengelompokan data untuk strategi promosi penerimaan mahasiswa baru, algoritma *K-means Clustering* adalah algoritma yang paling akurat dan algoritma yang paling mudah dipelajari untuk pengelompokan diantara algoritma lainnya, sehingga penulis memakai algoritma ini dalam melakukan pengelompokan data dengan melihat referensi beberapa penelitian sebelumnya.

2.5 Data Mining

Menurut Maburur A.G mengatakan Data *mining* adalah proses menganalisis data dari perspektif yang berbeda dan menggabungkannya menjadi informasi penting yang dapat digunakan untuk meningkatkan keuntungan, mengurangi biaya, atau bahkan keduanya. Secara teknis, data mining dapat digambarkan sebagai proses mencari korelasi atau pola diantara ratusan atau ribuan field dalam database relasional yang besar (Yunita, 2018).

Menurut Yunita (2018), Kemampuan data mining untuk menemukan informasi bisnis yang berharga dari database yang sangat besar dapat dibandingkan dengan menambang logam mulia dari negara asalnya, teknologi ini digunakan untuk :

1. Memprediksi tren dan karakteristik bisnis, sedangkan data mining mengotomatiskan penemuan informasi prediktif dari database besar.
2. Menemukan pola yang sebelumnya tidak diketahui, dimana data mining “memindai” database dan kemudian mengidentifikasi pola yang sebelumnya tersembunyi dalam satu langkah.

Menurut Jiawei Han & Micheline kamber dalam (Wulandari et al., 2021) mengatakan Data mining adalah proses menemukan pola yang juga menyediakan Informasi model deskriptif, dapat dimengerti dan prediktif untuk sejumlah data yang sangat besar. Menurut F. Gorunescu dalam (Wulandari et al., 2021) menyebutkan ada beberapa tahapan data mining antara lain :

1. Pembersihan Data
Sebaiknya data yang tidak relevan dibuang, karena keberadaannya nantinya dapat mempengaruhi kualitas atau keakuratan hasil data mining.
2. Integrasi data
Integrasi data harus dilakukan dengan hati-hati, karena kesalahan dalam integrasi data dapat menyebabkan hasil yang tidak jelas dan bahkan menyesatkan tindakan di masa mendatang.
3. Transformasi data
Transformasi dan pemilihan data ini nantinya juga menentukan kualitas hasil data mining, karena teknik data mining tertentu memiliki beberapa karakteristik yang bergantung pada langkah ini.
4. Aplikasi Teknik Data Mining
Aplikasi teknik data mining sendiri hanya merupakan bagian dari proses data mining.
5. Evaluasi pola yang ditemukan
Pada langkah ini, hasil teknik data mining dievaluasi dalam bentuk

pola tipikal atau model prediksi untuk mengevaluasi apakah hipotesis yang ada telah terpenuhi. Jika ternyata hasil yang diperoleh tidak sesuai dengan hipotesis, maka dapat dipilih beberapa opsi, seperti: Berikan umpan balik untuk meningkatkan proses data mining, mencoba teknik data mining lain yang lebih sesuai, atau terima hasil ini sebagai hasil yang tidak terduga yang mungkin berguna.

Menurut Larose mengatakan, data mining dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan, yaitu (Setiawan, 2017) :

1. Deskripsi

Deskripsi adalah mencari cara untuk menggambarkan pola dan tren yang ada dalam data. Deskripsi dari pola dan tren sering memberikan penjelasan yang mungkin untuk suatu pola atau tren tersebut.

2. Estimasi

Estimasi hampir sama dengan klasifikasi, kecuali bahwa variabel target dievaluasi secara numerik dan bukan secara kategoris. Model dibangun menggunakan full dataset yang memberikan nilai variabel target sebagai nilai prediktif.

3. Prediksi

Prediksi sebagian besar sama dengan klasifikasi dan estimasi, kecuali bahwa prediksi memprediksi nilai hasil di masa mendatang.

4. Klasifikasi

Klasifikasi memiliki tujuan variabel kategori, misalnya klasifikasi pendapatan dapat dibagi menjadi tiga kategori, yaitu berpenghasilan tinggi, menengah dan rendah.

5. Clustering

Clustering adalah pengelompokan atau fokus kumpulan data, observasi, dan pembentukan kategori dari item-item yang serupa. Cluster adalah kumpulan record yang mirip satu sama lain dan berbeda dari record di cluster lain. Algoritma clustering mencoba membagi seluruh data menjadi kelompok-kelompok yang sejenis (homogen), dimana kemiripan record satu kelompok memiliki nilai

maksimum, sedangkan kesamaan dengan record kelompok lain memiliki nilai minimum.

6. Asosiasi

Tugas asosiasi dalam data mining adalah menemukan atribut yang terjadi secara bersamaan. Dalam dunia bisnis, ini lebih umum disebut market basket analysis.

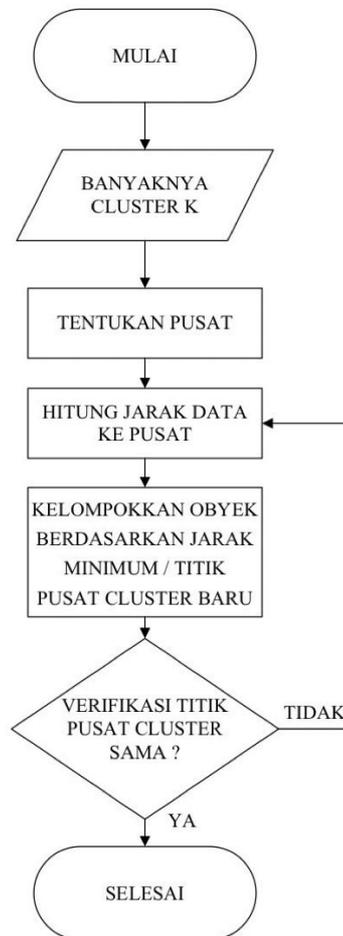
2.6 Clustering

Menurut Santosa dalam (Chasanah & Widiyono, 2017) mengatakan Pada dasarnya clustering adalah suatu metode pencarian dan pengelompokan data yang memiliki kesamaan karakteristik (similarity) antara data satu dengan lainnya. Clustering merupakan metode unsupervised data mining, artinya metode ini diterapkan tanpa pelatihan, tanpa pengajar, dan tidak memerlukan target output. Dalam data mining, dua jenis metode pengelompokan digunakan untuk mengelompokkan data, yaitu pengelompokan hierarkis dan pengelompokan non-hierarkis. Menurut Alfina et al., Ada dua metode clustering yang kita kenal, yaitu clustering hierarchical dan partitioning. Metode hierarchical clustering itu sendiri terdiri dari full linkage group, single linkage group, intermediate linkage group dan centroid linkage group, sedangkan metode partitioning sendiri terdiri dari k-means dan fuzzy k-means (Yunita, 2018).

2.7 Algoritma K-Means

K-Means adalah pembelajaran tanpa pengawasan (unsupervised) yang digunakan untuk mengelompokkan berdasarkan kesamaan karakteristik (Hairani et al., 2022). *K* berarti konstanta jumlah cluster yang diinginkan, *k-means* dalam hal ini berarti rata-rata dari suatu kelompok data yang dalam hal ini didefinisikan sebagai cluster, sehingga *K-means Clustering* adalah metode analisis data atau metode Data mining yang melakukan proses pemodelan tanpa pengawasan (Unsupervised) dan merupakan salah satu metode pengelompokan data menggunakan sistem partisi (Yunita, 2018). Menurut Susanto dalam (Mahmudi et

al., 2020) mengatakan Algoritma *K-means* pertama kali ditemukan oleh J. MacQueen dan berikut alur proses atau flowchart Algoritma *K-means Clustering* :



Gambar 2.5 Flowchart Algoritma *K-Means*

Algoritma *k-means* adalah algoritma yang membutuhkan parameter input sebanyak k dan membagi sekumpulan n objek kedalam k cluster sedemikian rupa sehingga kesamaan antara anggota cluster tinggi sedangkan kesamaan dengan anggota cluster lain sangat rendah (Yunita, 2018). Menurut Khotimah et al., Kemiripan anggota terhadap cluster diukur dari kedekatan objek terhadap nilai *mean* pada cluster atau dapat disebut sebagai centroid atau pusat cluster (Yunita, 2018).

Berikut adalah rumus untuk menentukan jarak data dari masing-masing centroid :

$$d(P, Q) = \sqrt{\sum_{j=1}^p (x_j(P) - x_j(Q))^2}$$

Keterangan :

D = titik dokumen

P = data *record*

Q = data *centroid*

Jarak yang terpendek antara centroid dengan dokumen menentukan posisi cluster suatu dokumen. Adapun rumus iterasi lainnya didefinisikan sebagai berikut :

$$C(i) = \frac{x_1 + x_2 + x_3 + \dots + x_n}{\sum x}$$

Keterangan :

X_1 = Nilai data *record* ke = 1

X_2 = Nilai data *record* ke = 2

$\sum x$ = jumlah data *record*

Suatu data akan menjadi anggota k-cluster jika jarak data dari pusat k-cluster ke-k lebih kecil dibandingkan dengan jarak ke pusat cluster lainnya (Chasanah & Widiyono, 2017). Menurut Langgeni, Baizal & W, Mengatakan Eksperimen ini menggunakan algoritma yang paling umum digunakan untuk clustering, yaitu algoritma k-means. Algoritma ini populer karena mudah diimplementasikan dan kompleksitas waktunya linier. Kerugiannya adalah algoritma ini sensitif terhadap inisialisasi cluster (Yunita, 2018). Dasar algoritmanya adalah sebagai berikut (Yunita, 2018) :

1. Inisialisasi kluster

2. Tempatkan setiap dokumen ke dalam cluster yang paling tepat berdasarkan ukuran kedekatan centroid. Centroid adalah vektor konseptual yang dianggap sebagai pusat cluster.
3. Setelah semua dokumen masuk ke dalam cluster. Hitung ulang fokus klaster berdasarkan dokumen klaster.
4. Jika centroid tidak berubah (dengan ambang batas tertentu), hentikan. Jika tidak, kembali ke langkah 2.

2.8 *Rapidminer*

Menurut Satria Wahono mengatakan Rapiminer merupakan aplikasi open source berlisensi AGPL (GNU Affero General Public License) yg bisa dipakai untuk mengolah data mining yg dikembangkan oleh Ralf Klinkenberg, Ingo Mierswa, dan Simon Fischer di Artificial Intelligence Unit dari University of Dortmund (Setiawan, 2017). Menurut F. A. Hendajani Mengatakan Tools *Rapidminer* digunakan untuk penyebaran karena mampu menganalisis data dalam jumlah yang sangat besar hingga tingkat data yang besar dengan visibilitas yang baik (Isnanto & Widodo, 2021).

2.9 *Davies Bouldin Index*

Davies Bouldin Index (DBI) merupakan salah satu metode yang diperkenalkan oleh David L. Davies dan Donald W. Bouldin. Menurut S.Narwin dalam (Adhitama et al., 2020) Davies Bouldin Index digunakan untuk mengevaluasi cluster secara umum berdasarkan kuantitas dan kedekatan antar anggota cluster semakin kecil nilai Davies Bouldin Index maka semakin baik cluster yang dihasilkan.

2.10 Strategi Promosi

Menurut Kotler & Armstrong mengatakan strategi promosi merupakan variabel variabel, di dalam *promotional mix* ada 5, yaitu (Setiawan, 2017) : Advertising, Personal selling, Sales promotion, Public relation dan Direct marketing.