

**ANALISIS CLUSTERING DENGAN K-MEANS
UNTUK PENGELOMPOKKAN PENJUALAN
PRODUK PADA HOTEL NEWTON**

SKRIPSI



**Oleh:
Event
171510008**

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNIK DAN KOMPUTER
UNIVERSITAS PUTERA BATAM
TAHUN 2021**

**ANALISIS CLUSTERING DENGAN K-MEANS
UNTUK PENGELOMPOKKAN PENJUALAN
PRODUK PADA HOTEL NEWTON**

SKRIPSI

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



**Oleh:
Event
171510008**

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNIK DAN KOMPUTER
UNIVERSITAS PUTERA BATAM
TAHUN 2021**

SURAT PERNYATAAN ORISINALITAS

Yang bertanda tangan di bawah ini saya:

Nama : Event
NPM : 171510008
Fakultas : Teknik dan Komputer
Program Studi : Sistem Informasi

Menyatakan bahwa “Skripsi” yang saya buat dengan judul:

Analisis *Clustering* dengan K-means untuk Pengelompokan Penjualan Produk pada Hotel Newton

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 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, 23 Januari 2021



Event
171510008

**ANALISIS CLUSTERING DENGAN K-MEANS
UNTUK PENGELOMPOKKAN PENJUALAN
PRODUK PADA HOTEL NEWTON**

SKRIPSI

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

**Oleh
Event
171510008**

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

Batam, 21 Januari 2021

A handwritten signature in black ink, consisting of several loops and a long horizontal stroke at the bottom.

a.n.

**Intan Utnasari, S.Kom., M.Kom.
Pembimbing**

ABSTRAK

Kepuasan pelanggan adalah tujuan utama yang ingin dicapai oleh semua perusahaan. Salah satu bidang usaha yang berfokuskan pada kepuasan pelanggan adalah bisnis perhotelan. Selain melayani akomodasi untuk pelanggan, hotel juga menyediakan beragam produk untuk dijual. Manajemen persediaan barang sangat penting dalam bisnis hotel karena salah satu cara untuk menjaga kepuasan pelanggan adalah dengan menyimpan stok agar tidak ada barang yang kosong pada saat dibutuhkan. Selain itu, manajemen persediaan barang yang baik tidak akan membuat perusahaan mengalami kerugian akibat persediaan yang sudah kadaluarsa. *Data mining* adalah metode umum untuk menyelesaikan masalah yang berkaitan dengan data. *Data mining* akan membantu dengan cara mengolah data yang telah ada dan menghasilkan pengetahuan baru terkait masalah tersebut yang diharapkan dapat menjadi solusi. Salah satu metode dalam data mining adalah *clustering* yang memiliki fungsi untuk mengelompokkan data dalam cluster sesuai kemiripannya. Algoritma K-means adalah algoritma yang paling umum digunakan dalam *data mining clustering*. Dengan menggunakan algoritma K-means *data mining clustering*, kita dapat mengelompokkan barang berdasarkan barang laku dan barang yang kurang laku. Penelitian ini bertujuan untuk membantu hotel Newton meningkatkan manajemen inventarisnya. Data yang digunakan adalah penjualan yang telah dilakukan yang dibagi menjadi 3 trimester. Data juga akan dievaluasi menggunakan aplikasi RapidMiner. Hasil yang diperoleh dari penelitian adalah 3 cluster dimana cluster tersebut terdiri dari barang sangat laku, barang sedang dan barang tidak laku.

Kata kunci : *data mining*, manajemen persediaan barang, algoritma K-means, *clustering*

ABSTRACT

Customer satisfaction is the main goal of many companies wants to achieve. One of the business fields that focus on customer satisfaction is hotel business. Apart from serving accommodation for customers, hotels also provide variety of products for sale. Inventory management is very important is hotel business because one of many ways to maintain customer satisfaction is by keeping stock so no items are empty when needed. In addition, a good inventory management will not let the company to experiences losses due to outdated inventory. Data mining is common method to solve problems related to datas. Data mining will help by processing late datas and make a new solution regarding the problems. One of the methods in data mining called clustering which put datas in clusters according to their similarity. K-means algorithm is the most common algorithm used in clustering data minig. Using data mining K-means clustering algorithm, we can group goods based on the salable goods and the less salable goods. This research aims to assist hotel Newton to improve their inventory management. Data that was used is sales that have been made which divided into 3 trimesters. Data also will be evaluated using RapidMiner application. The results obtained from the research are 3 clusters where the clusters consist of very salable goods, medium salable goods and not salable goods.

Keywords : data mining, inventory management, clustering, k-means algorithm

KATA PENGANTAR

Segala puji dan syukur penulis panjatkan ke hadirat-Nya yang telah memberikan segala rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan laporan tugas akhir yang merupakan salah satu persyaratan untuk menyelesaikan program studi strata satu (S1) pada Program Studi Sistem Informasi Universitas Putera Batam.

Penulis menyadari bahwa laporan akhir ini masih jauh dari kata sempurna, sehingga kritik dan saran yang sifatnya membantu penulis akan senantiasa penulis terima. Penulis menyadari bahwa laporan akhir ini takkan selesai tepat waktu tanpa bantuan, bimbingan, dan dorongan dari berbagai pihak. Untuk itu, pada kesempatan ini, penulis menyampaikan ucapan terima kasih kepada:

1. Kedua Orang Tua saya yang telah setia memberikan dorongan semangat, perhatian dan doa;
2. Ibu Dr. Nur Elfi Husda, S.Kom., M.SI. selaku Rektor Universitas Putera Batam;
3. Bapak Muhammad Rasid Ridho, S.Kom., M.SI. selaku Ketua Program Studi Sistem Informasi;
4. Ibu Intan Utnasari, S.Kom., M.Kom. selaku pembimbing Skripsi pada Program Studi Sistem Informasi Universitas Putera Batam;
5. Bapak Rika Harman, S.Kom., M.SI. selaku dosen pembimbing akademik;
6. Dosen dan Staff Universitas Putera Batam;
7. Teman-teman seperjuangan

Semoga kebaikan dalam membantu akan dibalaskan oleh-Nya dan selalu mendapatkan berkat-Nya. Amin.

Batam, 23 Januari 2021

Event

DAFTAR ISI

HALAMAN JUDUL.....	i
SURAT PERNYATAAN ORISINALITAS	iii
ABSTRAK	v
<i>ABSTRACT</i>	vi
KATA PENGANTAR	vii
DAFTAR ISI.....	viii
DAFTAR TABEL.....	x
DAFTAR GAMBAR	xi
DAFTAR RUMUS	xii
BAB I PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Identifikasi Masalah	4
1.3. Pembatasan Masalah	4
1.4. Rumusan Masalah	5
1.5. Tujuan Penelitian	5
1.6. Manfaat Penelitian	6
BAB II KAJIAN PUSTAKA	7
2.1. <i>Knowledge Discovery in Database (KDD)</i>	7
2.2. <i>Data Mining</i>	9
2.3. Metode <i>Data Mining</i>	13
2.4. <i>Software</i> Pendukung	16
2.5. Penelitian Terdahulu	18
2.6. Kerangka Pemikiran.....	22
2.7. Hipotesis Penelitian.....	23
BAB III METODE PENELITIAN.....	25
3.1. Desain Penelitian.....	25
3.2. Objek Penelitian	27
3.3. Populasi dan Sampel	28
3.4. Operasional Variabel.....	28
3.5. Teknik Pengumpulan Data.....	29

3.6	Model Penelitian	30
BAB IV HASIL DAN PEMBAHASAN		32
4.1	Hasil	32
4.2	Pembahasan.....	34
4.2.1	Seleksi Data.....	34
4.2.2	<i>Cleaning Data</i>	35
4.2.3	Perhitungan Algoritma	37
4.2.4	Evaluasi	42
BAB V SIMPULAN DAN SARAN		46
5.1	Simpulan	46
5.2	Saran.....	46
DAFTAR PUSTAKA		48
LAMPIRAN		49
	Lampiran 1. Daftar Riwayat Hidup.....	49
	Lampiran 2. Surat Izin Penelitian Universitas Putera Batam.....	50
	Lampiran 3. Surat Balasan Izin Penelitian Hotel Newton	51
	Lampiran 4. Foto Pendukung Penelitian	52
	Lampiran 5. Perhitungan Manual Iterasi ke-1.....	55
	Lampiran 6. Perhitungan Manual Iterasi ke-2.....	59
	Lampiran 7. Perhitungan Manual Iterasi ke-3.....	64

DAFTAR TABEL

Tabel 2.1 Tabel Penelitian Terdahulu	18
Tabel 3.1 Tabel Jadwal Penelitian	28
Tabel 4.1 Data Rekapitulasi Penjualan Produk Hotel Newton Tahun 2018	35
Tabel 4.2 Tabel data setelah seleksi data	36
Tabel 4.3 Tabel Hasil Iterasi ke-1	39
Tabel 4.4 Titik <i>centroid</i> baru hasil dari iterasi ke-1	40
Tabel 4.5 Tabel Hasil Iterasi ke-3	41
Tabel 4.6 Tabel Hasil Pengelompokkan Perhitungan Manual	44

DAFTAR GAMBAR

Gambar 2.1 Struktur Kerangka Pemikiran.....	23
Gambar 3.1 Desain Penelitian.....	25
Gambar 3.2 Struktur Organisasi Hotel Newton	27
Gambar 4.1 Tampilah Hasil <i>Folder View RapidMiner</i>	34
Gambar 4.2 Hasil Perhitungan <i>RapidMiner</i> Bentuk Grafik.....	42
Gambar 4.3 Hasil Perhitungan <i>RapidMiner</i> bentuk Deskripsi.....	43
Gambar 4.4 Hasil Perhitungan <i>RapidMiner</i> bentuk Tabel.....	44

DAFTAR RUMUS

Rumus 2.1 Rumus <i>Euclidean Distance</i>	15
Rumus 2.2 Rumus <i>centroid</i> baru.....	15
Rumus 3.1 Rumus <i>Euclidian Distance</i>	30
Rumus 4.1 Rumus <i>Euclidian Distance</i>	38

BAB I

PENDAHULUAN

1.1. Latar Belakang

Di era globalisasi, perkembangan teknologi sangatlah cepat. Perkembangan teknologi yang pesat ini merupakan salah satu aspek yang sepatutnya kita manfaatkan untuk mempermudah pekerjaan kita dalam sehari-hari, baik dalam bidang apapun. Kecanggihan teknologi yang pesat ditandai dengan semakin banyaknya penggunaan media digital sebagai alat bantu dalam kehidupan manusia. Penggunaan komputer dalam kehidupan sehari-hari juga menjadi salah satu bentuk perkembangan teknologi. Jika sebelum menggunakan komputer, penjualan suatu perusahaan akan direkap secara manual pada sebuah buku besar, maka sekarang komputer dapat membantu hal itu agar pekerjaan manusia semakin mudah. Penjualan yang telah terjadi dalam sebuah perusahaan dapat disimpan dalam bentuk digital dalam sebuah komputer. Penjualan yang dilakukan oleh sebuah perusahaan dapat disebut juga sebagai data. Data merupakan hal yang sangat sering dijumpai di kehidupan manusia sehari-hari, hanya saja manusia sering tidak memperhatikannya. Data dapat ditemukan baik dalam aspek pekerjaan ataupun saat berada di rumah. Dalam aspek pekerjaan, data dapat didapatkan di segala bidang pekerjaan. Data yang paling sederhana untuk diperhatikan adalah data penjualan atau data pembelian suatu kantor atau perusahaan. Segala bentuk pembelian dan penjualan yang terjadi dapat dijadikan data dan jika data diolah dengan baik, dapat membantu meringankan pekerjaan manusia. Hanya saja, tidak semua perusahaan dapat mengolah data ini dengan

baik. Data yang tidak diolah lagi akan mengakibatkan penumpukkan data yang tidak berguna. Jika diibaratkan pada sebuah komputer, data tersebut akan tersimpan dan akhirnya akan memenuhi *memory* dari komputer tersebut. Jika hal ini terjadi di dunia nyata, maka akan sangat disayangkan data yang sebanyak itu tersimpan dan tidak diolah dengan baik, sebab jika data diolah dengan baik, akan menghasilkan sebuah pengetahuan yang sangat menguntungkan oleh manusia.

Salah satu bidang usaha yang melakukan pembelian dan penjualan barang adalah bidang perhotelan. Dalam bidang perhotelan, banyak sekali transaksi jual beli yang terjadi, baik makanan, minuman, ataupun barang keperluan hotel lainnya. Dengan jumlah pelanggan yang sangat banyak dan pelanggan yang keluar masuk setiap harinya, tidak menutup kemungkinan bahwa penjualan yang terjadi dalam suatu hotel sangatlah banyak. Penjualan yang sangat banyak tentunya menghasilkan data dengan jumlah yang sangat besar. Data yang sangat besar jika tidak diolah dan hanya disimpan saja, akan mengakibatkan kasus penumpukkan data. Pada Hotel Newton, pendataan sudah memanfaatkan komputer, namun hanya berupa pendataan saja, tanpa adanya pengolahan data tambahan. Hal ini sama saja akan mengakibatkan penumpukkan data. Dengan memanfaatkan *data mining* dalam pengolahan data yang ada, diharapkan data penjualan yang banyak tersebut akan menghasilkan sebuah manfaat yang dapat mempermudah pekerjaan manusia. Salah satu manfaat yang akan terlihat jika data penjualan hotel Newton dilakukan pengolahan data dengan *data mining*, adalah pengelompokkan produk yang laris dan tidak laris. Hotel Newton sering mengalami kendala dimana persediaan barang yang dimiliki tidak sesuai dengan tingkat penjualan produk tersebut. Sering terjadinya kasus dimana persediaan barang kurang daripada

kebutuhan sehingga mengakibatkan kehabisan persediaan barang. Kasus yang serupa terjadi sebaliknya dimana persediaan barang yang terlalu banyak dibandingkan dengan penjualan barang sehingga persediaan barang menumpuk dan akhirnya melewati masa kadaluarsa. Dengan melakukan pengelompokan tersebut, maka masalah seperti kekurangan atau kelebihan stok barang yang dijual dapat dihindari. Pengolahan persediaan barang penjualan merupakan pengetahuan dasar dalam melakukan bisnis, yang sering disebut dengan *stock management*. Tanpa adanya pengolahan persediaan barang yang baik, seorang penjual tidak akan sanggup untuk memenuhi permintaan pelanggan disebabkan kurangnya persediaan barang yang diminta. Jika persediaan barang yang berlebihan dapat menyebabkan penurunan kualitas barang seiring dengan waktu yang lama akibat tidak digunakan dan untuk persediaan barang seperti makanan dan minuman dapat kadaluarsa. Cara untuk menangani masalah manajemen persediaan barang adalah dengan melakukan pengolahan historis transaksi yang pernah terjadi dan dengan memanfaatkan *data mining* untuk melakukan pengolahan historis transaksi tersebut, maka akan semakin cepat dan efisien. Dengan memanfaatkan *data mining* untuk memproses pengolahan data penjualan yang ada, diharapkan dapat membantu hotel Newton selaku lokasi penelitian untuk menghindari kasus manajemen persediaan barang seperti kekurangan ataupun kelebihan barang persediaan.

Dari latar belakang di atas, maka peneliti mengangkat judul berupa “Analisis Clustering dengan K-means untuk Pengelompokan Penjualan Produk pada Hotel Newton”

1.2. Identifikasi Masalah

Berdasarkan uraian latar belakang yang ada di atas, maka peneliti menyimpulkan permasalahan yang ada menjadi berikut :

1. Data penjualan belum dimanfaatkan secara maksimal
2. Membutuhkan waktu yang cukup lama untuk melakukan pengelompokan secara manual
3. Adanya kemungkinan terjadinya kasus kelebihan atau kekurangan stok barang yang dijual
4. Belum adanya cara yang efektif dan efisien yang dipraktekkan oleh pihak hotel.

1.3. Pembatasan Masalah

Untuk mencegah terjadi kesalahpahaman antar peneliti dengan pembaca serta untuk membatasi ruang lingkup penelitian yang sangat luas, peneliti memberikan batasan masalah dalam penelitian sebagai berikut :

1. Data yang diolah akan menggunakan metode K-means *clustering* dalam mengelompokkan data penjualan hotel
2. Data penelitian yang digunakan adalah data penjualan Hotel Newton.
3. Data yang akan digunakan dalam penelitian merupakan data penjualan hotel dalam *range* waktu dari bulan
4. *Software* yang akan digunakan dalam melakukan validasi adalah menggunakan *software RapidMiner*.

1.4. Rumusan Masalah

Berdasarkan identifikasi masalah yang ada di atas, maka peneliti merumuskan masalah yang ada sebagai berikut :

1. Bagaimana penerapan *data mining* dalam melakukan pengelompokkan data penjualan produk agar manajemen persediaan barang hotel Newton menjadi baik?
2. Bagaimana implementasi *data mining* pada pengolahan data penjualan produk hotel dengan menggunakan aplikasi *RapidMiner*?

1.5. Tujuan Penelitian

Dari rumusan masalah yang ada di atas, tujuan yang ingin dicapai oleh peneliti pada penelitian ini adalah sebagai berikut :

1. Untuk menerapkan *data mining* dalam melakukan pengelompokkan data penjualan produk sehingga dapat meningkatkan manajemen persediaan barang pada hotel Newton menjadi baik.
2. Untuk mengimplementasikan *data mining* dalam pengolahan data penjualan produk hotel Newton dengan menggunakan aplikasi *RapidMiner*.

1.6. Manfaat Penelitian

Adapun manfaat penelitian yang diharapkan oleh peneliti dari penelitian ini adalah sebagai berikut:

1. Bagi Peneliti

- a. Penelitian ini diharapkan dapat bermanfaat dan memberikan pengetahuan kepada peneliti sendiri serta peneliti selanjutnya yang akan menggunakan penelitian ini sebagai referensi penelitian yang sejenis di masa yang akan datang.
- b. Dalam pelaksanaan penelitian ini, diharapkan pengetahuan peneliti mengenai ilmu yang dibahas di dalam penelitian ini semakin dalam dan menambah wawasan peneliti.
- c. Penelitian ini diharapkan dapat menjadi dasar oleh peneliti sendiri dalam mewujudkan implementasinya secara *real* di dunia kerja nantinya.

2. Bagi Pihak Lain

- a. Diharapkan penelitian ini dapat membantu memudahkan pihak Hotel Newton untuk mengerti mengenai pentingnya pengolahan data dan *data mining* sehingga pihak Hotel Newton dapat mengimplementasikan pengolahan data dan meringankan pekerjaan bagi staff Hotel Newton.
- b. Untuk Universitas Putera Batam, diharapkan penelitian ini dapat menjadi referensi yang baik untuk peneliti selanjutnya yang akan menggunakan kasus atau permasalahan yang sejenis.

BAB II

KAJIAN PUSTAKA

2.1. *Knowledge Discovery in Database (KDD)*

Konsep *data mining* umumnya dikenal sebagai salah satu alat bantu yang sangat penting dalam manajemen pengolahan informasi. Hal ini disebabkan *data mining* dapat mengolah data yang sangat besar jumlahnya dan yang seperti yang kita ketahui data akan selalu bertambah. *Data mining* merupakan suatu langkah dalam *knowledge discovery in database*. *Knowledge discovery in database (KDD)* merupakan salah satu tahapan atau kegiatan yang meliputi kegiatan pengumpulan data dan diikuti pemakaian data historis untuk menemukan sebuah pola yang ada dalam hubungan antar data dalam sebuah *database* yang berukuran besar (K. Handoko, 2016). *Knowledge discovery in database* sering dikaitkan dengan *data mining* dikarenakan sering terjadinya kesalahpahaman tentang kedua istilah tersebut. Kedua istilah tersebut memiliki konsep yang berbeda, namun sangat berikatkan satu sama lain karena *data mining* merupakan salah satu tahapan dalam proses *knowledge discovery in database*. Dalam melakukan KDD, terdapat beberapa ciri penting dalam sebuah pencarian pengetahuan, yaitu (Aranski & Handoko, 2019):

1. Data yang dikerjakan dalam jumlah yang besar.
2. Volume data yang dikerjakan harus diefisiensikan.
3. Akurasi menjadi prioritas utama.
4. Menggunakan bahasa tingkat tinggi.
5. Memanfaatkan beberapa bentuk pembelajaran otomatis.

6. Hasil yang dihasilkan berupa pengetahuan yang bermanfaat.

Secara keseluruhan, proses KDD dapat dibagi menjadi beberapa tahapan yaitu sebagai berikut (K. Handoko & Lesmana, 2018):

1. *Data Selection* / Pemilihan Data

Sesuai dengan nama tahapan yang bersangkutan, pada tahap ini, proses pemilihan himpunan data dan menetapkan fokus akan suatu *database* terjadi di tahapan ini. Tahapan ini menjadi yang pertama dikarenakan harus adanya suatu *database* yang siap untuk dilakukan penggalian informasi sebelum proses *data mining* dimulai. Data hasil dari seleksi pada tahap ini akan digunakan dalam proses *data mining* nantinya.

2. *Data Cleaning* / Pembersihan Data

Pembersihan data merupakan salah satu tahapan yang paling perlu dilakukan pada sebuah data. Hal ini perlu dilakukan untuk menghindari proses *data mining* yang lama disebabkan oleh *noise*. *Noise* itu merupakan istilah untuk data yang tidak perlu dilakukan pemrosesan *data mining*. Bentuk-bentuk data yang tidak perlu diproses adalah data yang bersifat duplikasi, data yang inkonsisten serta data yang bersifat salah cetak (tipografi).

3. *Data Transformation* / Transformasi Data

Pencarian fitur-fitur yang berguna untuk mempresentasikan data bergantung kepada goal yang ingin dicapai. Merupakan proses transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses data mining. Proses ini merupakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam basis data.

4. *Data Mining*

Proses *Data mining* yaitu proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma dalam data mining sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan.

5. *Evaluation* / Evaluasi

Tahap ini merupakan bagian dari proses KDD yang mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesa yang ada sebelumnya.

2.2 *Data Mining*

Data mining merupakan suatu rangkaian proses untuk dimana *data mining* ini menggali nilai tambah dari sekumpulan data yang berupa pengetahuan yang selama ini tidak diketahui secara manual (K. Handoko, 2016). *Data mining* bukanlah suatu bidang yang sama sekali baru. Pada tahun 1990 istilah *data mining* mulai dikenal, ketika pemanfaatan data menjadi suatu yang penting dalam berbagai bidang, mulai dari bidang akademik, bisnis hingga medis. Munculnya *data mining* didasarkan pada jumlah data yang tersimpan dalam *database* yang semakin meningkat jumlahnya. Tujuan utama *data mining* adalah memanfaatkan data dalam basis data dengan mengolahnya sehingga menghasilkan informasi baru yang berguna (Iin Parlina, Agus Perdana Windarto, Anjar Wanto, 2018). *Data mining* sendiri sering disebut dengan istilah *knowledge discovery* karena tujuan

utama dari dilakukannya *data mining* adalah untuk mendapatkan sebuah pengetahuan yang baru (Suryadi, 2019).

Data mining bisa juga bisa memberikan dampak negatif maupun positif bergantung pada penggunaannya *data mining*nya. Hal ini terjadi jika pemanfaatan data yang dilakukan bersifat terlalu pribadi sehingga dapat dimanfaatkan oleh oknum yang tidak bertanggung jawab untuk memperoleh keuntungan lain yang bersifat merugikan pemilik data.

Data mining memiliki dua sudut pandang dalam pemanfaatannya, yaitu secara komersial dan keilmuan. Dari sudut pandang komersial, pemanfaatan *data mining* dapat digunakan untuk menangani meledaknya volume data dan dengan memanfaatkan *data mining*, diharapkan dapat menyelesaikan permasalahan ini.

Contoh pemanfaatan secara komersial yaitu:

1. Untuk mengetahui strategi pemasaran yang baik.
2. Untuk mengidentifikasi jenis barang yang harus dilakukan stok dalam jumlah yang tepat
3. Untuk memprediksi tingkat penjualan
4. Untuk memprediksi tingkat resiko produksi suatu item
5. Untuk mengidentifikasi karakteristik pelanggan.

Pada sudut pandang keilmuan *data mining* dapat digunakan untuk *capture*, menganalisis dan menyimpan data yang bersifat real time dan sangat besar. Misalnya, simulasi *scientific* yang akan menghasilkan data dalam ukuran *terabytes* atau lebih, pemanfaatan *remote sensor* dalam melakukan penempatan suatu satelit berdasarkan histori perputaran benda di luar angkasa.

Data mining memiliki tugas untuk memspesifikasi pola atau *pattern* yang ada dalam sebuah data. Dalam melakukan tugasnya, *data mining* memiliki kategori utama yang terdiri dari 2 yaitu (Vulandari, 2017):

1. Prediktif

Kategori *data mining* yang termasuk ke dalam prediktif memiliki tujuan untuk melakukan penebakan atau memprediksi nilai dari sebuah atribut tertentu berdasarkan nilai dari atribut tersebut. Pada kategori *data mining* yang prediktif terdapat variabel tak bebas yang disebut sebagai target dan atribut-atribut yang dimanfaatkan untuk membuat prediksi yang disebut *explanatory*.

2. Deskriptif

Data mining deskriptif bertujuan untuk menurunkan pola-pola yang meringkas hubungan pokok dalam data yang dilakukan proses *data mining*. Tugas *data mining* deskriptif lebih ke arah melakukan penyelidikan dan sering diikuti oleh proses *post-processing* untuk melakukan validasi serta penjelasan hasil yang dihasilkan.

Data mining memiliki beberapa metode berdasarkan jenis tugas-tugas yang akan dilakukan oleh *data mining* tersebut. Beberapa metode *data mining* yaitu (Hermawati, 2013):

1. *Estimation* / Estimasi

Estimasi merupakan proses klasifikasi namun dengan variabel target yang lebih mengarah ke arah numerik. Dimula dari memproses *record*

yang akan menjadi data dan diproses ke dalam proses *data mining* untuk melakukan estimasi.

2. *Prediction* / Prediksi

Prediksi merupakan proses melakukan tebakan di masa yang akan datang. Dengan menggunakan *data mining*, tebakan yang dilakukan berdasarkan hasil *data mining* dari data yang ada. Prediksi juga memanfaatkan klasifikasi ataupun estimasi dalam prakteknya.

3. *Classification* / Klasifikasi

Klasifikasi merupakan proses *data mining* yang bersifat menggolongkan data yang ada ke dalam variabel kategori yang ada. Metode klasifikasi memiliki tahapan yang hampir sama dengan pengelompokan.

4. *Clustering* / Pengelompokan

Clustering merupakan proses *data mining* yang bertujuan untuk mengelompokkan data yang diperoleh ke dalam suatu kelompok yang dimana data yang ada memiliki kemiripan antar satu data dengan yang lainnya dalam satu kelompok yang sama dan memiliki ketidakmiripan dengan data pada kelompok lainnya. *Clustering* hampir sama dengan klasifikasi, letak perbedaan terdapat pada kategori yang ada. Pada proses *clustering*, data tidak dimasukkan ke dalam penggolongan suatu kelompok dengan variabel target tertentu.

5. *Association* / Asosiasi

Asosiasi merupakan proses *data mining* yang bertujuan untuk menemukan suatu *attribute* yang muncul dalam suatu data dalam suatu

waktu. Bisa jadi data yang bertemu pada waktu secara bersamaan menimbulkan suatu atribut yang dapat ditemukan oleh *data mining*.

6. *Description* / Deskripsi

Deskripsi merupakan proses *data mining* yang bertujuan untuk secara sederhana menjelaskan penggambaran suatu pola yang muncul dari suatu *database* yang diolah menggunakan *data mining*.

2.3 Metode *Data Mining*

Pada penelitian ini, peneliti akan menggunakan metode *clustering* dalam penelitian ini. *Clustering* merupakan salah satu kategori dari *data mining* yang bersifat tanpa arahan / *unsupervised*. *Unsupervised* disini bermaksud tanpa adanya target kelompok yang disebutkan secara spesifik. *Clustering* merupakan salah satu dari banyak proses *data mining* yang tidak memerlukan latihan dan guru dalam melakukan proses *data mining* nya serta *clustering* tidak membutuhkan target *output* yang diinginkan secara spesifik. *Clustering* adalah proses pemisahan data ke dalam sejumlah kelompok dengan karakteristik yang mirip antara satu data dengan data lainnya pada kelompok yang sama. Target dari *clustering* adalah untuk melakukan pendistribusian data ke dalam kelompok-kelompok yang sedemikian sehingga derajat tingkat keterhubungan antar anggota kelompok yang sama kuat dan hubungan antar anggota kelompok yang berbeda lemah. *Clustering* tidak melakukan klasifikasi ataupun estimasi karena tidak adanya variabel target yang dimiliki oleh *clustering*. Algoritma *clustering* akan melakukan pembagian terhadap keseluruhan data dan menjadikannya ke dalam kelompok-kelompok yang *homogeny* (memiliki tingkat kemiripan maksimal) (K. Handoko, 2016).

Salah satu teknik yang digunakan dalam *clustering* adalah menggunakan algoritma *K-means*.

Algoritma *K-means* merupakan salah satu algoritma yang multivariat dengan tujuan untuk mengelompokkan objek-objek berdasarkan karakteristik yang dimilikinya. Algoritma *K-means* menggunakan *descriptive model* untuk menjelaskan algoritma pengelompokan yang dilakukannya (Putri et al., 2020). Algoritma *K-means* merupakan algoritma paling banyak digunakan dengan aplikasi penggunaan yang berukuran kecil hingga menengah. Hal ini terjadi karena kemudahan implementasinya serta memiliki waktu komputasi yang relatif cepat dan kemampuan algoritma dalam melakukan adaptasi terhadap kasus yang berbeda. Ide dasar algoritma *K-means* sangatlah sederhana, yaitu dengan meminimalkan *Sum of Squared Error* antara objek data dengan sejumlah k *centroid*. Namun kelemahan utama dari algoritma *K-means* adalah hasil yang terlalu sensitif terhadap pemilihan *cluster* awal dan perhitungan solusi lokal untuk mencapai kondisi optimal (Purnamasari & Dana, 2017).

Algoritma *K-means* terdiri dari beberapa langkah. Pertama, dari himpunan data yang akan dilakukan *clustering* dipilih sejumlah k objek secara acak yang akan menjadi *centroid* awal (k merupakan jumlah yang ditentukan oleh peneliti). Kedua, setiap objek yang bukan merupakan *centroid* awal dimasukkan ke dalam *cluster* terdekat. Ketiga, setiap *centroid* diperbarui berdasarkan rata-rata perhitungan dari objek yang ada dalam *cluster* tersebut. Perhitungan ini dilakukan pada setiap *cluster* sesuai jumlah yang ditentukan oleh peneliti. Keempat, ulangi langkah kedua dan ketiga dengan melakukan proses iterasi hingga semua *centroid* stabil. *Centroid* stabil berarti nilai *centroid* yang dihasilkan dalam iterasi yang saat

ini bernilai sama dengan nilai *centroid* yang dihasilkan dalam iterasi yang sebelumnya (Setiawan, 2018).

Jika dijabarkan secara rinci maka langkah-langkah algoritma *K-means* adalah sebagai berikut (S. Handoko et al., 2020):

1. Tentukan nilai k sebagai jumlah *cluster* yang diinginkan.
2. Tentukan nilai *centroid* awal secara acak.
3. Hitung jarak setiap data ke masing-masing *centroid* yang ada menggunakan rumus korelasi antar dua objek (*Euclidean Distance*)

$$(x_i, u_i) = \sqrt{(x_i - u_i)^2}$$

Sumber : (S. Handoko et al., 2020)

Rumus 2.1 Rumus *Euclidean Distance*

4. Kelompokkan setiap data berdasarkan jarak terdekat yang dihasilkan antara data tersebut dengan *centroid*.
5. Tentukan posisi *centroid* yang baru (k,C) dengan cara menghitung nilai rata-rata dari data yang ada pada *centroid* yang sama.

$$C_k = \left(\frac{1}{n_k}\right) \sum d_i$$

Sumber : (S. Handoko et al., 2020)

Rumus 2.2 Rumus *centroid* baru

6. Jika nilai *centroid* baru berbeda dengan nilai *centroid* lama, maka nilai *centroid* belum stabil. Lakukan lagi langkah ketiga hingga nilai *centroid*

yang dihasilkan memiliki nilai yang sama dengan nilai *centroid* yang ada pada proses iterasi sebelumnya

2.4 *Software* Pendukung

Software pendukung yang akan digunakan oleh peneliti pada penelitian ini adalah *software RapidMiner*. *RapidMiner* sebelumnya dikenal sebagai YALE (*Yet Another Learning Environment*). YALE dikembangkan pada tahun 2001 oleh sekelompok orang yang berasal dari Unit Kecerdasan Buatan Universitas Teknik Dortmund (*A.I Unit in TU Dortmund*). Kelompok tersebut terdiri dari Ralf Klinkenberg, Ingo Mierswa serta Simon Fisher. Dilanjutkan pada tahun 2006, Ralf dan Ingo mendirikan sebuah perusahaan yang bernama *Rapid-I*. Ralf dan Ingo melanjutkan untuk membangun serta mengembangkan YALE sehingga pada setahun kemudian pada tahun 2007, YALE mengalami perubahan nama menjadi *RapidMiner*. Pada tahun 2013, perusahaan dirian Ralf dan Ingo mengalami perubahan nama dari *Rapid-I* menjadi *RapidMiner* sesuai dengan nama aplikasi yang diciptakan oleh mereka. *RapidMiner* adalah *platform* perangkat lunak ilmu data yang dikembangkan oleh perusahaan yang memiliki nama yang sama. *RapidMiner* menyediakan lingkungan terintegrasi untuk melakukan persiapan data, *data mining*, AI, *word mining* guna mendukung penelitian, pendidikan, pelatihan, *rapid prototyping* serta pengembangan aplikasi. Aplikasi *RapidMiner* juga mendukung dalam melakukan proses pembelajaran mesin termasuk persiapan data, hasil visualisasi, validasi model dan optimasi data. Walaupun aplikasi *RapidMiner* sangat membantu dalam tugasnya, aplikasi ini masih bersifat *opensource* yang berarti model inti terbuka dan dapat didapatkan secara gratis

pada website yang disediakan dengan syarat penggunaan *RapidMiner* harus bertujuan dalam membantu pendidikan. Jika digunakan secara komersil, maka akan dikenakan biaya.

RapidMiner adalah salah satu *software* untuk pengolahan data mining. Pekerjaan yang dilakukan oleh *RapidMiner text mining* adalah berkisar dengan analisis teks, mengekstrak pola-pola dari data set yang besar dan mengkombinasikannya dengan metode statistika, kecerdasan buatan, dan database. Tujuan dari analisis teks ini adalah untuk mendapatkan informasi bermutu tertinggi dari teks yang diolah. *RapidMiner* menyediakan prosedur *data mining* dan *machine learning*, di dalamnya termasuk: ETL (*extraction, transformation, loading*), data *preprocessing*, *visualisasi*, *modelling* dan evaluasi. Proses data mining tersusun atas operator-operator yang *nestable*, dideskripsikan dengan XML, dan dibuat dengan GUI. Penyajiannya dituliskan dalam bahasa pemrograman *Java*. *RapidMiner* merupakan *software* / perangkat lunak untuk pengolahan data. Aplikasi *RapidMiner* sangat membantu dalam melakukan *data mining* karena di dalam aplikasi *RapidMiner* sudah terdapat prinsip serta algoritma yang sering digunakan dalam proses *data mining*, yang dimana salah satunya adalah algoritma *clustering K-means*. Dengan memanfaatkan algoritma yang telah dimilikinya, aplikasi *RapidMiner* dapat mengekstrak pola-pola dari *database* besar yang telah dimasukkan ke dalam aplikasi tersebut dengan mengkombinasikan metode statistika dan kecerdasan buatan yang dimilikinya. *RapidMiner* sangat membantu dalam memudahkan pengguna melakukan perhitungan data yang memiliki jumlah yang sangat banyak serta yang memiliki unsur operator yang banyak. Dalam penampilan hasil, aplikasi *RapidMiner* sangat

user-friendly dikarenakan hasil yang ditampilkan oleh aplikasi *RapidMiner* berupa visual dengan grafik yang ditemani oleh deskripsi penjelasan dari visual tersebut sehingga pengguna dapat mengerti dari visual yang ditampilkan.

2.5 Penelitian Terdahulu

Pada penelitian ini, peneliti menggunakan beberapa penelitian terdahulu yang pernah dilakukan oleh peneliti lainnya yang memiliki beberapa kemiripan dengan penelitian yang dilakukan oleh peneliti kali ini. Berikut merupakan penelitian terdahulu yang menjadi referensi oleh peneliti:

Tabel 2.1 Tabel Penelitian Terdahulu

No	Keterangan Jurnal	Judul	Pembahasan
1	Jurnal Penelitian Ilmu Komputer, <i>System Embedded & Logic</i> . Santoso Setiawan (2018) Volume 6, No. 1 p-ISSN : 2303-3304 e-ISSN : 2620-3553	Pemanfaatan Metode K-means dalam Penentuan Persediaan Barang	Membahas pengelompokan data produk dari historis transaksi dan menentukan produk yang membutuhkan stok paling banyak dan paling sedikit. Penelitian menggunakan 2 buah <i>cluster</i>
2	Jurnal Media Infotama Yulia Darmi dan Agus Setiawan	Penerapan <i>Clustering</i> K-means dalam	Membahas kasus manajemen persediaan barang yang kurang baik dan dengan implementasi <i>data mining</i> dapat membantu menyelesaikan permasalahan

	(2016) Volume 12, No. 2 ISSN : 1858-2680	Pengelompokan Penjualan Produk	tersebut. <i>Data mining</i> dilakukan dengan menggunakan 2 buah <i>cluster</i> .
3	Prosiding Seminar Nasional Geotik Herliyani Hasanah dan Wahyu Larasati (2019) ISSN : 2580-8796	Pemanfaatan Data Mining untuk Mengelompokkan Kategori Penjualan Produk	Jurnal membahas masalah persediaan barang yang laris dan tidak laris, pengelompokan dilakukan dengan mengelompokkan total produk ke dalam 3 buah <i>cluster</i> yang memiliki sifat laris hingga tidak laris dengan jumlah dan isi masing-masing kelompok berbeda
4	Jurnal Bianglala Informatika Elly Muningsih dan Sri Kiswati (2015) Volume 3, No. 1 ISSN : 2338-9761	Penerapan Metode K-means untuk Clustering Produk Online Shop dalam Penentuan Stok Barang	Membahas tentang manajemen persediaan barang di toko <i>online</i> yang tidak tepat, penelitian dilakukan berdasarkan data historis penjualan yang kemudian dikelompokkan menjadi 3 buah <i>cluster</i> dan peneliti menggunakan aplikasi <i>RapidMiner</i> untuk melakukan evaluasi pengolahan data yang dilakukan
5	Seminar Nasional Ilmu Sosial dan Teknologi 2 Alvendo Wahyu	Data Mining Clustering Nilai IQ Siswa	Jurnal membahas tentang masalah pengelompokan siswa yang kurang tepat dan peneliti dengan menggunakan nilai IQ sebagai tolak ukur pengelompokan.

	<p>Aranski dan Koko Handoko (2019)</p> <p>ISBN : 978-602-52829-1-1</p>	<p>Sekolah SMAK Yos Sudarso Batam</p>	<p>Penelitian menggunakan 3 buah <i>cluster</i> dalam melakukan pengelompokan dan menggunakan aplikasi <i>RapidMiner</i> untuk melakukan evaluasi pengolahan data yang dilakukan</p>
6	<p>Jurnal Teknologi dan Sistem Informasi</p> <p>Koko Handoko (2016)</p> <p>Volume 2, No. 3</p> <p>ISSN : 2476-8812</p>	<p>Penerapan Data Mining dalam Meningkatkan Mutu Pembelajaran pada Instansi Perguruan Tinggi Menggunakan Metode K-means Clustering</p>	<p>Membahas tentang masalah mutu pembelajaran yang dapat dikembangkan lagi pada sebuah instansi perguruan tinggi. Penelitian menggunakan <i>data mining</i> metode <i>clustering</i> K-means dan mengelompokan data ke dalam 4 buah <i>cluster</i>. Peneliti menggunakan aplikasi <i>RapidMiner</i> dalam evaluasi pengolahan data yang dilakukan</p>
7	<p>Seminar Nasional Ilmu Sosial dan Teknologi 1</p> <p>Koko Handoko dan Lido Sabda Lesmana (2018)</p>	<p>Data Mining pada Jumlah Penumpang Menggunakan Metode <i>Clustering</i></p>	<p>Penelitian membahas mengenai pengolahan data yang dilakukan pada bandara Hang Nadim dengan tujuan untuk mendapatkan jadwal padatnya jumlah penumpang pada bulan tertentu. Penelitian menggunakan metod K-means</p>

	ISBN : 978-602-52829-0-4		dan menggunakan aplikasi <i>RapidMiner</i> sebagai alat evaluasi pengolahan data.
8	Jurnal Sistem Informasi dan Komputer Resistiana Anggita Putri, Nida Inayah Maghfirani, et. Al (2020) Volume 9, No. 2 p-ISSN : 2301-7988 e-ISSN : 2581-0588	Analisis Pengelompokan Peraturan Kementerian dengan Menggunakan K-means Clustering	Penelitian membahas mengenai banyaknya peraturan kementerian Indonesia sehingga peneliti melakukan pengelompokan dengan tujuan untuk mengetahui permasalahan yang dibahas pada setiap peraturan. Penelitian menggunakan analisis <i>clustering</i> dengan memanfaatkan algoritma K-means. Dengan total data yang berjumlah 7233 peraturan, peneliti mengelompokan data ke dalam 4 buah <i>cluster</i> yang memiliki topik permasalahan masing-masing.
9	Informatika : Jurnal Ilmiah AMIK Labuhan Batu Sudi Suryadi (2019) Volume 6, No. 1 ISSN : 2303-2863	Penerapan Metode Clustering K-means untuk Pengelompokan Kelulusan Mahasiswa Berbasis	Penelitian membahas mengenai kesulitan yang dialami oleh salah satu sekolah tinggi yang dimana pihak akademik selalu menemui kesulitan dalam menentukan kandidat yang akan memenuhi permintaan mitra kerja dalam perekrutan karyawan. Penelitian menggunakan <i>data mining clustering</i> algoritma K-means dan data yang digunakan berupa nilai akademis dan

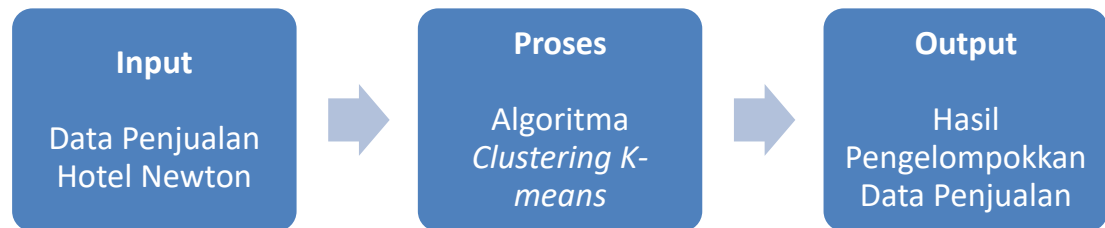
		Kompetensi	nilai kompetensi oleh mahasiswa yang dimiliki sekolah tinggi tersebut. Pada penelitian dilakukan pengelompokan data ke dalam 3 buah <i>cluster</i> yang dibagi berdasarkan nilai yang dimiliki setiap peserta.
10	<p>Indonesian Journal of Electrical Engineering and Computer Science</p> <p>Md. Zakir Hossain, Md. Nasim Akhtar, R.B. Ahmad, Mostafijur Rahman (2019)</p> <p>Volume 13, No. 2</p> <p>ISSN : 2502-4752</p>	<p><i>A Dynamic K-means Clusterin for Data Mining</i></p>	<p>Penelitian membahas tentang penggunaan algoritma <i>K-means</i> dengan jumlah <i>cluster</i> yang terlalu sedikit atau terlalu banyak dapat meningkatkan kemungkinan terjadinya pengelompokan yang kurang tepat sehingga penelitian menghasilkan algoritma <i>K-means</i> yang bersifat dinamis. Hasil dari penelitian menunjukkan perkembangan pesat yang dihasilkan oleh algoritma <i>K-means</i> dinamis yang dihasilkan oleh peneliti.</p>

Sumber : (Penulis, 2020)

2.6 Kerangka Pemikiran

Kerangka pemikiran dalam penelitian ini memiliki struktur input, proses dan output dalam satu jalan lurus. Input dalam penelitian ini berupa data penjualan Hotel Newton. Proses yang akan dilalui oleh input penelitian berupa algoritma *clustering K-means*. Output yang diharapkan dari proses penelitian berupa hasil

pengelompokkan penjualan. Jika digambarkan maka akan berbentuk seperti struktur di bawah ini:



Sumber : (Penulis, 2020)

Gambar 2.1 Struktur Kerangka Pemikiran

2.7 Hipotesis Penelitian

Hipotesis dalam penelitian ini adalah sebagai berikut:

1. a. Penerapan *data mining* pada pengelompokkan data penjualan produk membuat manajemen persediaan barang pada hotel Newton menjadi baik.
- b. Penerapan *data mining* pada pengelompokkan data penjualan produk membuat manajemen persediaan barang pada hotel Newton menjadi tidak baik.
2. a. Implementasi *data mining* dalam pengolahan data penjualan produk hotel Newton dengan menggunakan aplikasi *RapidMiner* berjalan dengan baik.

b. Implementasi *data mining* dalam pengolahan data penjualan produk hotel Newton dengan menggunakan aplikasi *RapidMiner* tidak berjalan dengan baik.

BAB III

METODE PENELITIAN

3.1 Desain Penelitian

Bentuk desain penelitian pada penelitian ini berbentuk sebagai berikut:



Sumber : (Penulis, 2020)

Gambar 3.1 Desain Penelitian

Berikut penjelasan masing-masing node pada desain penelitian yang ada di atas:

1. Identifikasi Masalah

Identifikasi masalah dimulai dari menentukan latar belakang sebuah permasalahan hingga menyimpulkan masalah tersebut menjadi pokok masalah dalam sebuah penelitian. Pada penelitian ini, permasalahan yang ada adalah data penjualan yang tidak diolah dengan menggunakan *data mining* yang mengakibatkan proses stok barang yang tidak efektif dan efisien.

2. Studi Pustaka

Studi pustaka yang dilakukan peneliti pada penelitian ini adalah riset dan pembelajaran mengenai permasalahan serta solusi yang diperlukan melalui referensi-referensi yang ada sehingga dapat menghasil metode yang baik dalam menyelesaikan permasalahan yang ada pada penelitian ini. Referensi yang digunakan antara lain buku, jurnal dan penelitian sebelumnya yang memiliki permasalahan atau memanfaatkan metode yang sama.

3. Analisa Data

Analisa data yang dilakukan peneliti pada penelitian ini adalah melakukan pengumpulan data primer dari lokasi penelitian yang telah ditentukan lalu dilakukan pembersihan pada data dari *noise* yang sering disebut *data cleaning*.

4. *Data Mining*

Data mining pada penelitian ini adalah melakukan proses perhitungan memanfaatkan algoritma *K-means*.

5. Evaluasi

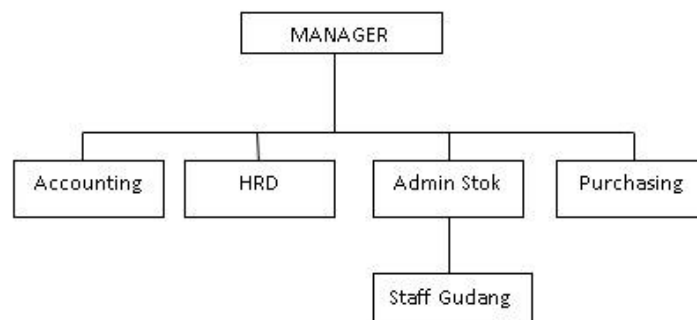
Evaluasi pada penelitian ini bertujuan untuk mengetahui ketepatan yang dihasilkan dari proses *data mining* yang telah dilakukan. Pada penelitian ini, evaluasi akan menggunakan alat bantu berupa aplikasi *RapidMiner*.

6. Hasil Penelitian

Setelah hasil dari *data mining* dilakukan evaluasi, maka akan didapatkan hasil penelitian. Pada tahapan ini, hasil penelitian akan dijelaskan agar dapat dimengerti oleh masyarakat.

3.2 Objek Penelitian

Objek penelitian berarti topik permasalahan yang dikaji dalam sebuah penelitian atau permasalahan yang diinvestigasi dalam sebuah penelitian. Dalam penelitian ini, objek penelitian yang digunakan adalah manajemen persediaan barang yang kurang tepat disebabkan data yang tidak diolah dengan efektif. Dalam penelitian ini, yang akan menjadi variabel adalah jenis barang yang dijual oleh hotel Newton serta jumlah yang terjual sesuai dengan hasil rekapitulasi tahunan hotel. Berikut merupakan struktur organisasi dari Hotel Newton selaku lokasi penelitian:



Sumber : (Penulis, 2020)

Gambar 3.2 Struktur Organisasi Hotel Newton

Dalam penelitian ini, yang menjadi lokasi penelitian adalah Hotel Newton. Hotel Newton terletak tepat di tengah daerah Nagoya Newton. Lokasi hotel Newton sangat strategis disebabkan lokasi tersebut sangat ramai dan sering dikunjungi oleh turis luar negeri khususnya turis Singapura dan Malaysia. Di sekitar hotel Newton juga banyak yang berjualan makan, baik pagi atau malam, sehingga tidak susah untuk mencari makanan.

Untuk waktu penelitian, penelitian dimulai dari akhir bulan September 2020 dan dijadwalkan untuk selesai pada Januari 2021. Berikut tabel jadwal penelitian yang direncanakan:

Tabel 3.1 Tabel Jadwal Penelitian

Kegiatan	Tahun, Bulan dan Pertemuan														
	September	Oktober					November			Desember				Januari	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	
Kajian Teori															
Metodologi Penelitian															
Pengumpulan Data															
<i>Data Mining</i>															
Kesimpulan															
Penyelesaian Akhir															

Sumber : (Penulis, 2020)

3.3 Populasi dan Sampel

Populasi dalam penelitian ini adalah data hasil rekapitulasi tahunan di tahun 2018. Namun, karena data yang dianggap terlalu banyak sehingga dapat memakan waktu yang lama dalam melakukannya, maka dari populasi tersebut ditarik sampel. Sampel yang digunakan adalah hasil rekapitulasi penjualan Hotel Newton tahun 2018 yang dibagi ke dalam bentuk trimester.

3.4 Operasional Variabel

Pada penelitian ini, operasional variabel yang akan digunakan dalam penelitian ini adalah barang produk yang dijual pada hotel Newton pada tahun 2018 dibagi ke dalam 3 trimester adalah sebagai berikut:

1. *Pocari Sweat*
2. *Coca-Cola*
3. Susu Beruang
4. *Red Bull*

5. *Sprite*
6. Lasegar Kaleng
7. *Carlsberg-CAN*
8. *Guinness Stout-CAN*
9. *Heineken-CAN*
10. GG Surya
11. *Marlboro Lights*
12. *Marlboro Menthol*
13. *Marlboro Red*
14. Sampoerna H
15. Sampoerna M
16. *Evolution Red*
17. *Evolution Menthol*
18. Tissue Kotak

3.5 Teknik Pengumpulan Data

Pada penelitian ini, peneliti melakukan beberapa teknik yang bertujuan untuk mengumpulkan data yang berhubungan dengan penelitian ini. Berikut teknik-teknik yang digunakan oleh peneliti:

1. Observasi

Peneliti melakukan observasi langsung ke lokasi penelitian. Tujuan observasi langsung dilakukan agar peneliti dapat merasakan langsung permasalahan yang ada serta untuk meminta izin dengan pihak manajemen untuk mendapatkan data primer yang digunakan.

2. Studi Pustaka

Studi pustaka dilakukan untuk mengetahui pengetahuan yang diperlukan untuk melakukan penelitian ini seperti metode yang paling tepat untuk dalam penelitian ini.

3.6 Model Penelitian

Model yang digunakan dalam penelitian ini adalah model kuantitatif. Model kuantitatif adalah model penelitian yang menguji teori tertentu dengan cara meneliti hubungan antar variabel yang ada dengan tolak ukur berupa angka dan dapat dianalisa dengan menggunakan prosedur-prosedur tertentu.

Prosedur yang digunakan pada penelitian ini adalah menggunakan algoritma *data mining K-means* yang bertujuan untuk mengelompokkan data-data yang memiliki sifat yang mirip antar satu data dengan data lainnya. Pemilihan algoritma yang digunakan dikarenakan algoritma *K-means* merupakan algoritma dengan tujuan pengelompokkan yang dapat dilakukan dengan cepat dan efektif untuk data yang berskala tidak terlalu banyak. Algoritma *K-means* menggunakan proses iterasi perhitungan untuk mendapatkan suatu pola yang akan menjadi hasil dari algoritma tersebut. Rumus yang akan digunakan pada algoritma adalah rumus *Euclidian Distance* guna mengukur jarak antar titik dengan titik pusat. Rumus *Euclidian Distance* adalah sebagai berikut:

$$d(xi, ui) = \sqrt{(xi - ui)^2}$$

Sumber : (S. Handoko et al., 2020)

Rumus 3.1 Rumus *Euclidian Distance*

Setelah mengetahui data yang letaknya berdekatan, maka akan dilakukan pengelompokkan sesuai dengan jarak data yang mendekati dengan titik pusat, lalu dilakukan iterasi perhitungan ulang. Jika pola yang didapatkan telah sama, maka iterasi selesai dan pengelompokkan yang terjadi merupakan hasil dari algoritma.