

Implementasi Algoritma Apriori

dalam Toko *SparePart* Mandiri Cool Jaya

Oleh:

Nama : Nathanael

NIM : 57180298

Skripsi

Diajukan sebagai salah satu syarat

untuk memperoleh gelar Sarjana Komputer

Program Studi Teknik Informatika



KWIK KIAN GIE
SCHOOL OF BUSINESS

INSTITUT BISNIS dan INFORMATIKA KWIK KIAN GIE

JAKARTA

FEBRUARI 2022

© Hak cipta milik IBI KKG (Institut Bisnis dan Informatika Kwik Kian Gie)

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik dan tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IBIKKG.
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IBIKKG.



KWIK KIAN GIE
SCHOOL OF BUSINESS

Institut Bisnis dan Informatika Kwik Kian Gie

DAFTAR ISI

	DAFTAR ISI	Error! Bookmark not defined.
	BAB I PENDAHULUAN	4
	A. Latar Belakang Masalah	4
	B. Identifikasi Masalah.....	6
	C. Batasan Masalah	6
	D. Tujuan Penelitian	6
	E. Manfaat Penelitian	7
	BAB II LANDASAN TEORI.....	8
	A. <i>Business Analytics</i>	8
	B. <i>Data Mining</i>	9
	1. Proses Analitik Data	9
	2. Strategi <i>Data mining</i>	11
	3. Teknik <i>Data Mining</i> Dasar	14
	4. Evaluasi Performa.....	22
	C. <i>Association Rules (Aturan Asosiasi)</i>	24
	I. <i>Confidence & Support</i>	25
	D. Apriori.....	28
	E. Penelitian Sebelumnya.....	30
	BAB III ANALISIS SISTEM YANG BERJALAN	32

© Hak Cipta milik IBI KKG (Institut Bisnis dan Informatika Kwik Kian Gie)



1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik dan tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IBIKKG.
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IBIKKG.



A.	Gambaran Umum Objek Penelitian	32
B.	Analisis Sistem yang Berjalan	33
C.	Metodologi Penelitian	34
1.	Teknik Pengumpulan Data.....	34
2.	Teknik Analisis Data	34
3.	Teknik Pengukuran Data	35
D.	Jadwal Penelitian	36
	DAFTAR PUSTAKA.....	37

© Hak cipta milik IBI KKG (Institut Bisnis dan Informatika Kwik Kian Gie)
Hak Cipta Dilindungi Undang-Undang

Institut Bisnis dan Informatika Kwik Kian Gie

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik dan tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IBIKKG.
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IBIKKG.

BAB I

PENDAHULUAN

A. Latar Belakang Masalah

Manusia sebagai makhluk sosial memiliki arti bahwa manusia membutuhkan manusia lain. Dalam menjalankan kehidupan sehari-hari, manusia tidak dapat menjalankan hidupnya sendiri. Bahkan dalam memenuhi kebutuhannya, manusia memerlukan bantuan untuk menjalani kehidupan mereka. Hal ini berlaku untuk semua manusia. Setiap manusia selalu membutuhkan manusia lainnya. Setiap manusia dalam bermasyarakat pasti melakukan komunikasi, sosialisasi dan juga interaksi dengan masyarakat lainnya. Manusia sebagai makhluk sosial sudah terjadi sejak ia lahir. Seorang manusia yang akan lahir pun membutuhkan manusia lain untuk memberikan pertolongan.

Selain itu manusia juga membutuhkan bantuan alat-alat untuk memudahkan kehidupan mereka setiap harinya. Manusia dan teknologi adalah dua kata yang tak bisa dipisahkan pada era modern pada saat ini. Kita tahu bahwa perkembangan teknologi yang sangat pesat saat ini, seakan-akan memudahkan dan memanjakan manusia dalam menjalankan aktivitas kehidupannya. Hubungan manusia dan teknologi ini bisa ditemui di setiap aktivitas yang kita lakukan, Manusia tak akan pernah lepas dari sebuah Teknologi. Sehingga Teknologi saat ini, merupakan bagian dari kebutuhan pokok yang harus wajib dirasakan dan dinikmati manfaatnya, selain kebutuhan pokok seperti pangan, sandang, dan papan.

Iklim Indonesia secara keseluruhan adalah iklim tropis. Perairan yang hangat di wilayah Indonesia sangat berperan dalam menjaga suhu di darat tetap konstan, dengan



1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik dan tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IBIKKG.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IBIKKG.



suhu yang hangat sampai panas di wilayah pedalaman dan dataran tinggi, serta 23 °C di wilayah pegunungan. Perubahan suhu antar musim di Indonesia pun tidak begitu signifikan. Dikarenakan suhu yang hangat sampai panas di Indonesia maka teknologi pendingin seperti pendingin ruangan (*air conditioner / AC*) dan kulkas sangatlah dibutuhkan, baik untuk mendinginkan suhu ruangan yang panas maupun menjaga kesegaran makanan maupun minuman agar tetap segar.

Peralatan maupun teknologi juga membutuhkan *maintenance* agar tetap bekerja dan tidak mudah rusak. jadi sangatlah penting untuk merawat benda-benda tersebut yang membantu kita setiap harinya. Pada saat *maintenance* sangatlah penting juga untuk mengganti suku cadang (*SparePart*) untuk menjaga kualitas dari benda dan teknologi tersebut.

Perkembangan dunia usaha mengalami pertumbuhan yang pesat, hal ini menyebabkan persaingan antar perusahaan semakin kompetitif. Pertumbuhan dunia usaha yang pesat dikarenakan oleh berbagai faktor, salah satu contoh faktor yang mempengaruhi yaitu faktor teknologi yang dari zaman ke zaman terus berkembang semakin canggih. Perusahaan menyadari munculnya berbagai tantangan yang harus dihadapi, baik perusahaan sejenis di dalam negeri maupun yang berasal dari luar negeri.

Banyaknya toko *SparePart* di Indonesia saat ini mengalami perkembangan yang sangat pesat sehingga keadaan ini menimbulkan persaingan bisnis yang ketat dan kompetitif. Jumlah penduduk sekitar 250 juta jiwa, menjadikan Indonesia pasar yang menjanjikan bagi perusahaan *SparePart*. Mayoritas Toko *SparePart* menargetkan konsumen maupun perusahaan yang merakit maupun memperbaiki barang elektronik,

Kondisi tersebut juga membuat persaingan juga semakin tajam antar toko yang sejenis di Indonesia. Pihak penjual dituntut untuk jeli dalam menggali informasi



mengenai preferensi konsumen dan mampu mencari tahu barang dagangan apa saja yang harus diperhatikan *supply*-nya, dengan mengetahui barang apa yang sering dicari oleh konsumen dan penjual juga harus mengisi *supply* barang-barang yang laku tersebut agar dapat menyaingi toko lain dengan informasi yang didapatkan.

B. Identifikasi Masalah

Berdasarkan latar belakang di atas, maka masalah penelitian dapat diidentifikasi antara lain sebagai berikut :

1. Pihak penjual tidak begitu tahu mengenai preferensi konsumen.
2. belum tahu barang jualan apa saja yang harus diperhatikan *supply*nya,
3. penjual tidak tahu barang-barang apa saja yang sering terjual dan harus di isi *supply*nya.

C. Batasan Masalah

Berdasarkan latar belakang diatas Batasan masalahnya adalah

1. Belum adanya sistem untuk mengetahui preferensi pembeli
2. Belum adanya sistem untuk memberitahu barang apa saja yang harus diperhatikan *supply*nya.
3. Belum adanya sistem untuk memberitahu barang apa saja yang sering terjual dan harus diisi stoknya.

D. Tujuan Penelitian

Tujuan dari penelitian ini adalah untuk membantu penjualan di toko *SparePart* Mandiri Cool Jaya dengan menggunakan *data mining* agar dapat meningkatkan *sales* di



toko tersebut dengan mencari tahu preferensi pembeli, memperhatikan *supply* barang, dan mencari tahu barang apa saja yang sering terjual.

E. Manfaat Penelitian

Manfaat dari penelitian ini adalah :

1. Hakiptamilik IBI KKG (Institut Bisnis dan Informatika Kwik Kian Gie)

Hak Cipta Dilindungi Undang-Undang

1. Bagi perusahaan

- a. Mempermudah toko/pengusaha dalam menjual barang - barang dagangan yang laku.
- b. Meningkatkan penjualan di toko tersebut.

2. Bagi peneliti

- a. Mengetahui seberapa dibutuhkannya *apriori* untuk perusahaan.
- b. Mengetahui efisiensi dan efektivitas dari *apriori*.

3. Bagi pembaca

- a. Memberikan ide penggunaan *apriori* untuk membantu bisnis penjualan.

BAB II

Landasan Teori

A. *Business Analytics*

Menurut Galit, S. et al (2018:3) “*Business Analytics* (BA) adalah praktik dan seni membawa data kuantitatif untuk digunakan dalam pengambilan keputusan. Istilah ini memiliki arti yang berbeda untuk organisasi yang berbeda.”

Business Analytics, mencakup berbagai metode analisis data. Banyak aplikasi yang kuat melibatkan sedikit lebih dari menghitung, memeriksa aturan, dan aritmatika dasar. Untuk beberapa organisasi, inilah yang dimaksud dengan *analytics*.

Tingkat berikutnya dari analitik bisnis, sekarang disebut *Business Intelligence* (BI), mengacu pada visualisasi data dan pelaporan untuk memahami "apa yang terjadi dan apa yang terjadi." Ini dilakukan dengan menggunakan bagan, tabel, dan dasbor untuk menampilkan, memeriksa, dan menjelajahi data. BI, yang sebelumnya sebagian besar terdiri dari pembuatan laporan statis, telah berkembang menjadi alat dan praktik yang lebih ramah pengguna dan efektif, seperti membuat dasbor interaktif yang memungkinkan pengguna tidak hanya mengakses data waktu nyata tetapi juga berinteraksi langsung dengannya. Dasbor yang efektif adalah yang mengikat langsung ke data perusahaan dan memberi manajer alat untuk melihat dengan cepat apa yang mungkin tidak mudah terlihat dalam database kompleks yang besar. Salah satu alat untuk manajer operasi industri menampilkan pesanan pelanggan dalam satu tampilan dua dimensi, menggunakan warna dan ukuran gelembung sebagai variabel tambahan, menunjukkan nama pelanggan, jenis produk, ukuran pesanan, dan lama waktu untuk memproduksi.



1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik dan tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IBIKKG.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IBIKKG.



B. Data Mining

Menurut Richard, J.R. (2017:5) “*data mining* didefinisikan sebagai proses menemukan struktur yang menarik dalam data. Struktur dapat mengambil banyak bentuk, termasuk seperangkat aturan, grafik atau jaringan, pohon, satu atau beberapa persamaan, dan banyak lagi. Strukturnya dapat menjadi bagian dari dasbor visual yang kompleks atau sederhana daftar kandidat politik dan nomor terkait yang mewakili sentimen pemilih berdasarkan umpan Twitter.”

Menurut Parteek Bhatia (2019:17) “Data mining adalah kumpulan teknik untuk penemuan otomatis yang efisien dari pola yang sebelumnya tidak diketahui, valid, baru, berguna, dan dapat dipahami dalam database besar. Pola harus dapat ditindaklanjuti sehingga dapat digunakan dalam pengambilan keputusan perusahaan.”

Menurut Galit, S. et al (2018:5) “*Data Mining* mengacu pada metode analisis bisnis yang melampaui hitungan, teknik deskriptif, pelaporan, dan metode berdasarkan aturan bisnis.”

Sesi *Data Mining* menggunakan satu atau beberapa algoritme untuk tujuan mengidentifikasi tren dan pola yang menarik dalam data. Pengetahuan yang diperoleh dari sesi *data mining* adalah model umum dari data. Tujuan utamanya adalah untuk menerapkan apa yang telah ditemukan pada situasi baru. Analisis Data adalah suatu proses atau upaya pengolahan data menjadi sebuah informasi baru agar karakteristik data tersebut menjadi lebih mudah dimengerti dan berguna untuk solusi suatu permasalahan, khususnya yang berhubungan dengan penelitian.

1. Proses Analitik Data:

a. Mendapatkan Data

Data mining membutuhkan akses ke data. Data dapat direpresentasikan sebagai *volume record* dalam beberapa file *database*, atau data mungkin hanya berisi beberapa ratus *record* dalam satu *file*. Kesalahpahaman yang umum adalah bahwa untuk membangun model yang



efektif, algoritma *data mining* harus disajikan dengan ribuan atau jutaan contoh. Faktanya, sebagian besar alat penambangan data bekerja paling baik dengan beberapa ratus atau beberapa ribu catatan terkait. Oleh karena itu, setelah masalah didefinisikan, langkah pertama dalam proses *data mining* adalah mengekstrak atau merakit *subset data* yang relevan untuk diproses. Sering kali, langkah pertama ini membutuhkan banyak waktu dan usaha manusia.

Praproses Data

Pra-pemrosesan data adalah bagian proses analitik yang paling memakan waktu dan paling tidak menguntungkan. Data dapat berisi nilai dan noise yang hilang dan memerlukan satu atau beberapa transformasi sebelum siap untuk ditambang. Satu tugas prapemrosesan dasar melibatkan pemeriksaan tingkat korelasi antara pasangan atribut input numerik

Data Mining

Seperti Namanya, proses ini menambang data dari data yang di ekstrak dan di proses sebelumnya.

Interprestasi dan Mengevaluasi Data

Interpretasi hasil mengharuskan kami untuk memeriksa keluaran alat penambangan data untuk menentukan apakah hasil yang telah ditemukan bermanfaat dan menarik. jika hasilnya kurang optimal, kita dapat mengulangi langkah *data mining* menggunakan atribut dan/atau *instance* baru. Atau kembali ke data *warehouse* dan mengulangi proses ekstraksi data.

Laporkan/terapkan Hasil

Tujuan utamanya adalah menerapkan apa yang telah ditemukan pada situasi baru. Misalkan melalui proses analisis pasar *data mining*, peneliti

c. Hak cipta milik IBI KKG (Institut Bisnis dan Informatika Kwik Kian Gie)

Hak Cipta Dilindungi Undang-Undang

d. Institut Bisnis dan Informatika Kwik Kian Gie

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik dan tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IBIKKG.
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IBIKKG.



menemukan bahwa produk X hampir selalu dibeli dengan produk Y. Contoh klasiknya adalah penemuan bahwa persentase yang sangat tinggi dari orang yang membeli popok bayi pada hari Kamis juga membeli bir . Reaksi kejutan awal terhadap temuan ini masuk akal ketika kita menyadari bahwa pasangan dengan bayi kecil di rumah kemungkinan tidak akan pergi keluar pada Jumat atau Sabtu malam, melainkan lebih memilih untuk menikmati akhir pekan dengan bersantai di rumah. Seorang analis pasar dapat memanfaatkan temuan ini dengan menjadikan bir sebagai barang pajangan yang jelas bagi pelanggan yang membeli popok.

© Hak cipta milik IBI KKG (Institut Bisnis dan Informatika Kwik Kian Gie)

Hak Cipta Dilindungi Undang-Undang

2. Strategi *Data mining*

Strategi *data mining* dapat secara luas diklasifikasikan sebagai diawasi atau tidak diawasi. Pembelajaran yang diawasi membangun model dengan menggunakan atribut *input* untuk memprediksi nilai atribut *output*. Banyak algoritma penambangan data yang diawasi hanya mengizinkan atribut keluaran tunggal. Alat pembelajaran terawasi lainnya memungkinkan kita untuk menentukan satu atau beberapa atribut keluaran. Atribut *output* juga dikenal sebagai variabel dependen karena hasilnya bergantung pada nilai dari satu atau lebih atribut *input*. Atribut *input* disebut sebagai variabel independen. Ketika pembelajaran tidak diawasi, atribut keluaran tidak ada. Oleh karena itu, semua atribut yang digunakan untuk membangun model adalah variabel bebas.

Strategi dalam *data mining* ada 5 yaitu :

- a. Klasifikasi

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik dan tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IBIKKG.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IBIKKG.



Klasifikasi mungkin yang paling dipahami dari semua strategi *data mining*. Tugas klasifikasi memiliki tiga karakteristik umum yaitu Pembelajaran diawasi, Variabel terikat bersifat kategoris.

Beberapa contoh tugas klasifikasi antara lain sebagai berikut:

- (1) Tentukan karakteristik yang membedakan individu yang pernah mengalami serangan jantung dengan yang tidak.
- (2) Kembangkan profil orang yang "sukses".
- (3) Mengklasifikasikan pemohon kredit mobil sebagai risiko kredit yang baik atau buruk.
- (4) Menentukan apakah pengajuan pajak penghasilan telah dipalsukan

Estimasi

Seperti klasifikasi, tujuan dari model estimasi adalah untuk menentukan nilai untuk atribut keluaran yang tidak diketahui. Namun, tidak seperti klasifikasi, atribut keluaran untuk masalah estimasi adalah numerik daripada kategoris. Berikut adalah empat contoh tugas estimasi:

- (1) Perkirakan jumlah menit sebelum badai petir mencapai lokasi tertentu.
- (2) Perkirakan laba atas investasi untuk kampanye iklan.
- (3) Perkirakan panjang ledakan sinar gamma

Prediksi

Tidak mudah membedakan prediksi dari klasifikasi atau estimasi. Namun, tidak seperti model klasifikasi atau estimasi, tujuan model prediktif adalah untuk menentukan hasil masa depan daripada perilaku saat ini. Atribut keluaran dari model prediktif dapat berupa kategorikal atau numerik. Berikut adalah beberapa contoh tugas yang sesuai untuk penambangan data prediktif:

c. Hak cipta milik IBI KKG (Institut Bisnis dan Informatika Kwik Kian Gie)

Hak Cipta Dilindungi Undang-Undang

Institut Bisnis dan Informatika Kwik Kian Gie

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik dan tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IBIKKG.
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IBIKKG.



- (1) Memprediksi jumlah total *touchdown* yang akan dicetak oleh NFL selama musim NFL 2017
- (2) Tentukan apakah pelanggan kartu kredit kemungkinan akan memanfaatkan penawaran khusus yang tersedia dengan tagihan kartu kreditnya
- (3) Prediksi harga penutupan *Dow Jones Industrial Average* minggu depan

d. *Unsupervised Clustering*

Dengan pengelompokan tanpa pengawasan, kami tanpa variabel terikat untuk memandu proses pembelajaran. Sebaliknya, program pembelajaran membangun struktur pengetahuan dengan menggunakan beberapa ukuran kualitas *cluster* untuk mengelompokkan contoh ke dalam dua atau lebih kelas. Tujuan utama dari strategi pengelompokan tanpa pengawasan adalah untuk menemukan struktur konsep dalam data.

Penggunaan umum dari pengelompokan tanpa pengawasan meliputi:

1. Mendeteksi penipuan dalam aktivitas perdagangan saham, klaim asuransi, atau transaksi keuangan
2. Menentukan apakah hubungan yang bermakna dalam bentuk konsep dapat ditemukan dalam data yang mewakili kilatan ledakan sinar gamma di luar tata surya kita
3. Menentukan apakah data yang tersedia untuk umum tentang bank yang gagal berguna untuk membangun model yang diawasi untuk memprediksi kegagalan bank

e. analisis keranjang pasar

c. Hak cipta milik IBI KKG (Institut Bisnis dan Informatika Kwik Kian Gie)

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik dan tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IBIKKG.
2. Dilarang menggunakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IBIKKG.



Tujuan dari analisis keranjang pasar adalah untuk menemukan hubungan yang menarik antara produk ritel. Hasil analisis keranjang pasar membantu pengecer merancang promosi, mengatur rak atau barang katalog, dan mengembangkan strategi pemasaran silang. Algoritme aturan asosiasi, sering digunakan untuk menerapkan analisis keranjang pasar ke sekumpulan data.

3 Teknik *Data Mining* Dasar

Hak Cipta Dilindungi Undang-Undang

Institut Bisnis dan Informatika Kwik Kian Gie

a. *Decision Tree*

a. *Decision Tree*

Decision Tree adalah struktur populer untuk pembelajaran yang diawasi. Banyak artikel telah ditulis tentang keberhasilan penerapan model *Decision Tree* untuk masalah dunia nyata.

Decision Tree dibangun hanya dengan menggunakan atribut-atribut yang paling mampu membedakan konsep yang akan dipelajari. Sebuah *Decision Tree* dibangun dengan awalnya memilih *subset* dari contoh dari set pelatihan. *Subset* ini kemudian digunakan oleh algoritma untuk membangun *Decision Tree*. *Instance* set pelatihan yang tersisa menguji keakuratan pohon yang dibangun. Jika *Decision Tree* mengklasifikasikan contoh dengan benar, prosedur berakhir. Jika sebuah *instance* salah diklasifikasikan, *instance* tersebut ditambahkan ke *subset* yang dipilih dari *instance* pelatihan, dan pohon baru akan dibangun. Proses ini berlanjut hingga pohon yang mengklasifikasikan semua *instance* yang tidak dipilih dengan benar dibuat atau *Decision Tree* dibuat dari seluruh rangkaian pelatihan. Kami menawarkan versi algoritme yang disederhanakan yang menggunakan

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik dan tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar IBIKKG.

2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IBIKKG.



seluruh rangkaian *instance* pelatihan untuk membangun pohon keputusan. Langkah-langkah algoritmanya adalah sebagai berikut:

C Hak cipta milik IBI KKG (Institut Bisnis dan Informatika Kwik Kian Gie)

Hak Cipta Dilindungi Undang-Undang

Institut Bisnis dan Informatika Kwik Kian Gie

- (1) Misalkan T adalah himpunan *instance* pelatihan.
- (2) Pilih atribut yang paling membedakan *instance* yang terkandung dalam T.
- (3) Buat simpul pohon yang nilainya adalah atribut yang dipilih. Buat *link* anak dari simpul ini dimana setiap tautan mewakili nilai unik untuk atribut yang dipilih. Gunakan nilai *link* anak untuk membagi lebih lanjut *instance* ke dalam subkelas.

(4) Untuk setiap *subclass* yang dibuat pada langkah 3,

- (a) sebuah. Jika *instance* dalam *subclass* memenuhi kriteria yang telah ditentukan sebelumnya atau jika kumpulan pilihan atribut yang tersisa untuk jalur pohon ini adalah nol, tentukan klasifikasi untuk *instance* baru yang mengikuti jalur keputusan ini.
- (b) Jika subkelas tidak memenuhi kriteria yang telah ditentukan sebelumnya dan setidaknya ada satu atribut untuk membagi jalur pohon lebih lanjut, misalkan T adalah himpunan *instance* subkelas saat ini dan kembali ke langkah 2

Pilihan atribut yang dibuat saat membangun *Decision Tree* menentukan ukuran pohon yang dibangun. Tujuan utamanya adalah untuk meminimalkan jumlah level pohon dan simpul pohon, sehingga memaksimalkan generalisasi data.

Decision Tree memiliki beberapa keunggulan. Berikut adalah daftar beberapa dari banyak keuntungan yang ditawarkan pohon keputusan.

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik dan tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IBIKKG.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IBIKKG.



Hak Cipta Dilindungi Undang-Undang

- (1) *Decision Tree* mudah dipahami dan dipetakan dengan baik ke seperangkat aturan produksi.
- (2) *Decision Tree* telah berhasil diterapkan pada masalah nyata.
- (3) *Decision Tree* tidak membuat asumsi sebelumnya tentang sifat data.
- (4) *Decision Tree* mampu membangun model dengan kumpulan data yang berisi data numerik dan juga kategorikal

Seperti semua algoritma *data mining*, ada beberapa masalah seputar penggunaan pohon keputusan. Secara khusus,

- (1) Atribut keluaran harus kategorikal, dan beberapa atribut keluaran tidak diperbolehkan.
- (2) Algoritme *Decision Tree* tidak stabil karena sedikit variasi dalam data pelatihan dapat menghasilkan pemilihan atribut yang berbeda pada setiap titik pilihan di dalam pohon. Efeknya bisa signifikan karena pilihan atribut memengaruhi semua subpohon turunan.
- (3) Pohon yang dibuat dari kumpulan data numerik bisa sangat kompleks karena pemisahan atribut untuk data numerik biasanya biner

b. *Basic Covering Rule*

Ingat bahwa sebuah *instance* dikatakan tercakup oleh aturan jika memenuhi prasyarat aturan. Untuk setiap kelas konsep, tujuannya adalah untuk membuat seperangkat aturan yang memaksimalkan jumlah total *instance* dalam kelas yang tercakup sementara pada saat yang sama meminimalkan jumlah *instance* nonkelas yang tercakup. Di sini diuraikan teknik penutup langsung berdasarkan algoritma PRISM Metode dimulai dengan aturan tunggal yang kondisi antesedennya kosong dan konsekuennya

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik dan tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar IBIKKG.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IBIKKG.

diberikan sebagai kelas yang mewakili *instance* yang akan dicakup. Pada titik ini, domain dari semua *instance* kumpulan data dicakup oleh aturan. Artinya, ketika aturan dieksekusi, itu mengklasifikasikan setiap *instance* sebagai anggota kelas yang ditentukan dalam konsekuensi aturan. Untuk membatasi cakupan aturan ke *instance-instance* di dalam kelas, nilai atribut input dianggap satu per satu sebagai kemungkinan tambahan untuk anteseden aturan. Untuk setiap iterasi, kombinasi atribut-nilai yang dipilih adalah yang paling mewakili *instance* tersebut di dalam kelas. Aturan ditambahkan ke daftar aturan penutup setelah prasyarat aturan dipenuhi hanya oleh anggota kelas yang bersangkutan atau daftar atribut telah habis. Setelah aturan ditambahkan, *instance* yang dicakup oleh aturan akan dihapus dari kumpulan *instance* yang tidak ditemukan. Proses berlanjut sampai semua *instance* dari kelas yang bersangkutan dicakup oleh satu atau lebih aturan. Metode ini kemudian diulang untuk semua kelas yang tersisa. Beberapa variasi dari teknik umum ini ada, tetapi konsep penutupnya sama

c. *Association Rules*

Analisis afinitas adalah proses umum untuk menentukan hal-hal mana yang berjalan bersama. analisis keranjang pasar, dimana keinginannya adalah untuk menentukan barang-barang yang kemungkinan akan dibeli oleh pelanggan selama pengalaman berbelanja. Keluaran dari analisis keranjang pasar adalah sekumpulan asosiasi tentang perilaku pembelian pelanggan. Asosiasi diberikan dalam bentuk seperangkat aturan khusus yang dikenal sebagai aturan asosiasi. Aturan asosiasi digunakan untuk membantu menentukan strategi pemasaran produk yang tepat. Di bagian ini, kami menjelaskan prosedur yang efisien untuk menghasilkan aturan asosiasi.

© Hak cipta milik IBI KKG (Institut Bisnis dan Informatika Kwik Kian Gie)

Hak Cipta Dilindungi Undang-Undang

Institut Bisnis dan Informatika Kwik Kian Gie



1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik dan tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IBIKKG.
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IBIKKG.



d. *K-Means*

Algoritma *K-means* adalah teknik pengelompokan statistik yang sederhana namun efektif untuk data numerik. Untuk membantu Anda lebih memahami pengelompokan tanpa pengawasan, mari kita lihat bagaimana algoritme *K-means* mempartisi sekumpulan data menjadi kluster yang terpisah. Berikut algoritmanya:

- (1) Pilih nilai untuk *K*, jumlah total *cluster* yang akan ditentukan.
- (2) Pilih *K instance* (titik data) dalam kumpulan data secara acak. Ini adalah pusat *cluster* awal.
- (3) Gunakan jarak *Euclidean* sederhana untuk menetapkan *instance* yang tersisa ke pusat *cluster* terdekat.
- (4) Gunakan *instance* di setiap *cluster* untuk menghitung *mean* baru untuk setiap *cluster*.
- (5) Jika nilai *mean* yang baru identik dengan nilai *mean* dari iterasi sebelumnya, proses dihentikan. Jika tidak, gunakan cara baru sebagai pusat *cluster* dan ulangi langkah 3-5.

Langkah pertama dari algoritme membutuhkan keputusan awal tentang berapa banyak *cluster* yang kami yakini ada dalam data. Selanjutnya, algoritma secara acak memilih *K* titik data sebagai pusat *cluster* awal. Setiap *instance* kemudian ditempatkan di *cluster* yang paling mirip. Kesamaan dapat didefinisikan dalam banyak cara; Namun, ukuran kesamaan yang paling sering digunakan adalah jarak *Euclidean* sederhana. Setelah semua *instance* ditempatkan di *cluster* yang sesuai, pusat *cluster* diperbarui dengan menghitung rata-rata setiap *cluster* baru. Proses klasifikasi *instance* dan komputasi pusat *cluster* berlanjut hingga iterasi algoritma menunjukkan tidak

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik dan tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar IBIKKG.

2. Dilarang menggunakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IBIKKG.

ada perubahan pada pusat *cluster*, yaitu sampai tidak ada *instance* yang mengubah penetapan *cluster* pada langkah 3.

Metode *K-means* mudah dipahami dan diterapkan. Namun, ada beberapa masalah yang perlu dipertimbangkan. Secara khusus,

1. Algoritme hanya bekerja dengan data bernilai nyata. Jika kita memiliki atribut kategorikal dalam kumpulan data kita, kita harus membuang atribut tersebut atau mengubah nilai atribut menjadi ekuivalen numerik. Pendekatan umum adalah membuat satu atribut numerik untuk setiap nilai atribut kategorikal.
2. Kita diharuskan untuk memilih nilai jumlah *cluster* yang akan dibentuk. Ini adalah masalah yang jelas jika kita membuat pilihan yang buruk. Salah satu cara untuk mengatasi masalah ini adalah dengan menjalankan algoritme beberapa kali dengan nilai alternatif untuk *K*. Dengan cara ini, kita lebih mungkin untuk mendapatkan "rasa" tentang berapa banyak *cluster* yang mungkin ada dalam data.
3. Algoritma *K-means* bekerja paling baik ketika *cluster* yang ada dalam data berukuran kira-kira sama. Dalam kasus ini, jika solusi optimal diwakili oleh *cluster* dengan ukuran yang tidak sama, algoritma *K-means* tidak mungkin menemukan solusi terbaik.
4. Tidak ada cara untuk mengetahui atribut mana yang signifikan dalam menentukan *cluster* yang terbentuk. Untuk alasan ini, beberapa atribut yang tidak relevan dapat menyebabkan hasil yang kurang optimal.
5. Kurangnya penjelasan tentang sifat kelompok yang terbentuk membuat kita bertanggung jawab atas banyak interpretasi tentang apa yang telah ditemukan. Namun, seperti yang akan Anda lihat di bab

C Hak cipta milik IBI KKG (Institut Bisnis dan Informatika Kwik Kian Gie)

Hak Cipta Dilindungi Undang-Undang

Institut Bisnis dan Informatika Kwik Kian Gie



1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik dan tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IBIKKG.
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IBIKKG.

selanjutnya, kita dapat menggunakan alat penambangan data yang diawasi untuk membantu kita mendapatkan wawasan tentang sifat kluster yang dibentuk oleh algoritme pengelompokan yang tidak diawasi. Terlepas dari keterbatasan ini, algoritma *K-means* terus menjadi teknik statistik favorit.

e. *Genetic Learning*

Algoritma genetika menerapkan pendekatan evolusioner untuk pembelajaran induktif. Model pembelajaran algoritma genetika pada awalnya dikembangkan oleh John Holland pada tahun 1986 dan didasarkan pada prinsip seleksi alam Darwinian. Pemrograman genetik telah berhasil diterapkan pada masalah yang sulit dipecahkan dengan menggunakan teknik konvensional. Area aplikasi yang umum termasuk masalah penjadwalan, seperti masalah wiraniaga keliling, masalah perutean jaringan untuk jaringan *circuit-switched*, dan masalah di bidang pemasaran keuangan. Algoritma genetika dapat dikembangkan untuk penambangan data yang diawasi dan tidak diawasi. Berikut adalah algoritma pembelajaran genetik dasar:

- (1) Inisialisasi populasi P dari n elemen, sering disebut sebagai kromosom, sebagai solusi potensial.
- (2) Sampai kondisi terminasi tertentu terpenuhi,
 - (a) Gunakan fungsi *fitness* untuk mengevaluasi setiap elemen dari solusi saat ini. Jika sebuah elemen melewati kriteria *fitness*, elemen tersebut tetap berada di P ; jika tidak, itu dibuang dari populasi.

© Hak cipta milik IBI KKG (Institut Bisnis dan Informatika Kwik Kian Gie)

Hak Cipta Dilindungi Undang-Undang



1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik dan tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IBIKKG.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IBIKKG.

- (b) Populasi sekarang berisi m elemen ($m \leq n$). Gunakan operator genetik untuk membuat $(n - m)$ elemen baru. Tambahkan elemen baru ke populasi untuk mengembalikan ukuran populasi ke n .

Untuk penambahan data, kami menganggap elemen sebagai *instance* yang ditentukan oleh atribut dan nilai. Teknik umum untuk merepresentasikan pengetahuan genetik adalah dengan mengubah elemen menjadi *string biner*.

Operator genetik yang paling luas adalah *crossover* dan mutasi. *Crossover* membentuk elemen baru untuk populasi dengan menggabungkan bagian dari dua elemen yang saat ini ada dalam populasi. Elemen yang paling sering digunakan untuk *crossover* adalah elemen yang ditakdirkan untuk dihilangkan dari populasi. Mutasi, operator genetik kedua, sedikit diterapkan pada elemen yang dipilih untuk dieliminasi. Mutasi dapat diterapkan dengan membalik bit (atau nilai atribut) secara acak dalam satu elemen. Atau, pilihan apakah akan membalik bit individu dapat didasarkan pada probabilitas. Seleksi adalah operator genetik ketiga yang kadang-kadang digunakan. Dengan seleksi, elemen yang dihapus dari populasi diganti dengan salinan elemen yang lulus uji *fitness* dengan skor tinggi. Dengan cara ini, kebugaran keseluruhan populasi dijamin meningkat. Karena seleksi mendukung spesialisasi daripada generalisasi, itu harus digunakan dengan hati-hati.

Algoritme genetika terus mendapatkan popularitas, dan beberapa produk penambahan data komersial sekarang mengandung komponen pembelajaran genetik. Berikut adalah daftar pertimbangan ketika menggunakan pendekatan pemecahan masalah berdasarkan pembelajaran genetik:

© Hak cipta milik IBI KKG (Institut Bisnis dan Informatika Kwik Kian Gie)

Hak Cipta Dilindungi Undang-Undang

Institut Bisnis dan Informatika Kwik Kian Gie



1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik dan tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IBIKKG.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IBIKKG.



- (1) Algoritme genetika dirancang untuk menemukan solusi yang dioptimalkan secara global. Namun, tidak ada jaminan bahwa solusi apa pun yang diberikan bukanlah hasil dari optimasi lokal dan bukan global.
- (2) Fungsi *fitness* menentukan kompleksitas komputasi dari algoritma genetika. Fungsi kebugaran yang melibatkan beberapa perhitungan dapat menjadi mahal secara komputasi.
- (3) Algoritme genetika menjelaskan hasil mereka sejauh fungsi *fitness* dapat dimengerti.
- (4) Mengubah data ke bentuk yang cocok untuk algoritme genetika dapat menjadi tantangan.

© Hak cipta milik IBI KKG (Institut Bisnis dan Informatika Kwik Kian Gie)

Hak Cipta Dilindungi Undang-Undang

4 Evaluasi Performa

Evaluasi kinerja mungkin yang paling penting dari semua langkah dalam proses *data mining*. Di bagian ini, kami menawarkan pendekatan akal sehat untuk mengevaluasi model pembelajar terawasi dan tidak terawasi. Dalam bab-bab selanjutnya, kita akan berkonsentrasi pada teknik evaluasi yang lebih formal. Sebagai titik awal, kami mengajukan tiga pertanyaan umum:

- a. Apakah manfaat yang diterima dari proyek *data mining* lebih dari mengimbangi biaya proses *data mining*?
- b. Bagaimana kita menginterpretasikan hasil dari sesi *data mining*?
- c. Bisakah kita menggunakan hasil proses *data mining* dengan percaya diri?



Ketiga pertanyaan tersebut sulit untuk dijawab. Namun, yang pertama lebih merupakan tantangan karena beberapa faktor ikut bermain. Berikut adalah daftar minimal pertimbangan untuk pertanyaan pertama:

- a. Apakah ada pengetahuan tentang proyek sebelumnya yang serupa dengan proyek yang diusulkan? Berapa tingkat keberhasilan dan biaya proyek yang serupa dengan proyek yang direncanakan?
- b. Apa bentuk data yang akan dianalisis saat ini? Apakah datanya ada, atau harus dikumpulkan? Ketika banyak data ada dan tidak dalam bentuk yang dapat diterima untuk penambahan data, biaya proyek terbesar akan termasuk dalam kategori persiapan data. Sebenarnya, pertanyaan yang lebih besar mungkin adalah apakah akan mengembangkan gudang data untuk proyek penambahan data di masa depan.
- c. Siapa yang akan bertanggung jawab atas proyek *data mining*? Berapa banyak karyawan saat ini yang akan terlibat? Apakah konsultan luar akan dipekerjakan?
- d. Apakah perangkat lunak yang diperlukan saat ini tersedia? Jika tidak, apakah perangkat lunak akan dibeli atau dikembangkan? Jika dibeli atau dikembangkan, bagaimana perangkat lunak akan diintegrasikan ke dalam sistem saat ini?

Seperti yang Anda lihat, setiap jawaban untuk pertanyaan pertama membutuhkan pengetahuan tentang model bisnis, status data yang tersedia saat ini, dan sumber daya saat ini. Oleh karena itu, kami akan mengalihkan perhatian kami untuk menyediakan alat evaluasi untuk pertanyaan 2 dan 3. Kami pertama-tama mempertimbangkan evaluasi model pembelajar terawasi.

Hak Cipta Dilindungi Undang-Undang

Hak cipta milik IBI KKG (Institut Bisnis dan Informatika Kwik Kian Gie)

Institut Bisnis dan Informatika Kwik Kian Gie

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik dan tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar IBIKKG.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IBIKKG.



C. Association Rules (Aturan Asosiasi)

Menurut Richard, J.R. (2017:47) “*association rule mining* digunakan untuk menemukan asosiasi yang menarik antara atribut-atribut yang terdapat dalam sebuah *database*.”

Tidak seperti aturan produksi tradisional, aturan asosiasi dapat memiliki satu atau beberapa atribut keluaran. Juga, atribut *output* untuk satu aturan dapat menjadi atribut *input* untuk aturan lain. Aturan asosiasi adalah teknik populer untuk analisis keranjang pasar karena semua kemungkinan kombinasi pengelompokan produk yang berpotensi menarik dapat dieksplorasi. Untuk alasan ini, sejumlah atribut yang terbatas dapat menghasilkan ratusan aturan asosiasi. Analisis afinitas adalah proses umum untuk menentukan hal-hal mana yang berjalan bersama. Aplikasi tipikal adalah analisis keranjang pasar, dimana keinginannya adalah untuk menentukan barang-barang yang kemungkinan akan dibeli oleh pelanggan selama pengalaman berbelanja. Keluaran dari analisis keranjang pasar adalah sekumpulan asosiasi tentang perilaku pembelian pelanggan. Asosiasi diberikan dalam bentuk seperangkat aturan khusus yang dikenal sebagai aturan asosiasi. Aturan asosiasi digunakan untuk membantu menentukan strategi pemasaran produk yang tepat.

Aturan asosiasi berusaha mengidentifikasi kombinasi hal-hal yang sering terjadi bersama (analisis afinitas). Ini juga merupakan dasar dari analisis keranjang pasar, yang kami diskusikan dalam hal korelasi dan rasio Jaccard. Aturan asosiasi mengambil langkah lebih jauh dengan menerapkan bentuk pembelajaran mesin, yang paling umum adalah algoritma apriori.

Ada banyak kegunaan dari aturan asosiasi. Secara klasik, mereka diterapkan pada analisis transaksi ritel, mirip dengan analisis keranjang pasar. Dengan munculnya data

Hak Cipta Dilindungi Undang-Undang
Hak Cipta milik IBIKK (Institut Bisnis dan Informatika Kwik Kian Gie)

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik dan tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IBIKKG.
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IBIKKG.



aturan ke atribut tunggal. Generator aturan asosiasi memungkinkan konsekuensi aturan berisi satu atau beberapa nilai atribut. Untuk menunjukkan hal ini, misalkan kita ingin menentukan apakah ada hubungan menarik yang dapat ditemukan dalam tren pembelian pelanggan di antara empat produk toko kelontong berikut: Susu, Keju, Roti, Telur.

Kemungkinan asosiasi termasuk yang berikut adalah

- (1) Jika pelanggan membeli susu, mereka juga membeli roti.
- (2) Jika pelanggan membeli roti, mereka juga membeli susu.
- (3) Jika pelanggan membeli susu dan telur, mereka juga membeli keju dan roti.
- (4) Jika pelanggan membeli susu, keju, dan telur, mereka juga membeli roti

Asosiasi pertama memberi tahu kita bahwa pelanggan yang membeli susu juga cenderung membeli roti. Pertanyaan yang jelas adalah, “Seberapa besar kemungkinan peristiwa pembelian susu mengarah pada pembelian roti?” Untuk menjawab ini, setiap aturan asosiasi memiliki kepercayaan yang terkait. Untuk aturan ini, kepercayaan diri adalah probabilitas bersyarat dari pembelian roti yang diberikan pembelian susu. Oleh karena itu, jika total 10.000 transaksi pelanggan melibatkan pembelian susu, dan 5000 transaksi yang sama juga berisi pembelian roti, kepercayaan pembelian roti dengan pembelian susu adalah $5000/10.000 = 50\%$.

Sekarang pertimbangkan aturan kedua. Apakah aturan ini memberi kita informasi yang sama seperti yang pertama? Jawabannya adalah tidak! Dengan aturan pertama, domain transaksi terdiri dari semua pelanggan yang telah melakukan pembelian susu. Untuk aturan ini, domain adalah kumpulan semua transaksi pelanggan yang menunjukkan pembelian item roti. Sebagai contoh, misalkan kita memiliki total 20.000

Hak Cipta Dilindungi Undang-Undang

Hak cipta milik IBI KKG (Institut Bisnis dan Informatika Kwik Kian Gie)

Institut bisnis dan Informatika Kwik Kian Gie

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik dan tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar IBIKKG.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IBIKKG.



set item. Langkah kedua menggunakan set item yang dihasilkan untuk membuat seperangkat aturan asosiasi.

Ketika beberapa atribut hadir, pembuatan aturan asosiasi menjadi tidak masuk akal karena banyaknya kemungkinan kondisi untuk konsekuensi dari setiap aturan. Algoritma khusus telah dikembangkan untuk menghasilkan aturan asosiasi secara efisien. Salah satu algoritma tersebut adalah algoritma Apriori. Algoritma ini menghasilkan apa yang dikenal sebagai *itemset*. Kumpulan item adalah kombinasi nilai atribut yang memenuhi persyaratan cakupan yang ditentukan. Kombinasi nilai atribut yang tidak memenuhi persyaratan cakupan akan dibuang. Hal ini memungkinkan proses pembuatan aturan diselesaikan dalam waktu yang wajar.

Pembuatan aturan asosiasi Apriori menggunakan set item adalah proses dua langkah. Langkah pertama adalah pembuatan set item. Langkah kedua menggunakan set item yang dihasilkan untuk membuat seperangkat aturan asosiasi.

Algoritma apriori beroperasi dengan menemukan semua aturan yang memenuhi *confidence* minimum dan spesifikasi *support*. Contoh Algoritma untuk menentukan kandidat *itemset* C_k dalam ukuran K :

1. For $K = 1$ menghasilkan semua *itemset* dengan $support \geq Support_{min}$
if *itemset* Null, berhenti dan tingkatkan K sebanyak 1,
jika *itemset* dari K sudah diidentifikasi semua dengan $support \geq Support_{min}$,
END(mengidentifikasi *frequent item* L1).
2. Return daftar dari *frequent itemset*.
3. Mengidentifikasi *rules* dalam bentuk *antecedents* dan *consequents* dari *frequent item*.



4. Mengecek *confidence* dari *rules*.

Jika *confidence* suatu *rules* memenuhi $confidence_{min}$, maka *rules* tersebut menjadi kuat kuat.

Output dari algoritma apriori dapat digunakan sebagai dasar untuk merekomendasikan juga mempertimbangkan faktor-faktor seperti korelasi, atau analisis dari teknik lain, dari kumpulan *dataset*. Algoritma apriori dapat menghasilkan banyak *frequent itemset*. Aturan asosiasi dapat dibuat dengan hanya melihat *frequent itemset* yang kuat, dalam arti memenuhi atau melampaui level *support* minimum dan tingkat *confidence* minimum.

E. Penelitian Sebelumnya

Dalam penulisan skripsi ini penelitian terdahulu menjadi salah satu bahan acuan penulis untuk melakukan penelitian sehingga penulis dapat memperkaya teori dalam mengkaji penelitian yang dilakukan dari penelitian terdahulu, adapun beberapa penelitian terdahulu yang berkaitan dengan penelitian yang dilakukan yaitu :

1. J. Chandra dan K.R. Dewi (2020) pada jurnal nya yang berjudul “Implementation of Data Mining Sales of Milk Using Apriori Algorithm Method” pada penelitian yang dilakukan yaitu membahas implementasi data mining pada penjualan susu untuk mempermudah dalam melakukan stok bahan, mengumpulkan informasi konsumen dengan mudah.
2. O. Chouat dan A.H. Irawan. (2020) yang berjudul “Implementation of Data Mining on Online Shop in Indonesia” yang menjelaskan tentang toko online di Indonesia yang diimplementasikan dengan data mining. Pada implementasi yang dilakukan yaitu bertujuan untuk memudahkan

pengumpulan informasi data dari pengguna sehingga nanti dapat diolah untuk sistem yang akan dibuat dan bisa membantu pada stok barang, penjualan, dan menganalisis data pembeli dan pemilik.

Erma, D.S. (2018), Penerapan Data Mining Penjualan Sepatu Menggunakan Metode Algoritma Apriori, penelitian ini membahas tentang penerapan *data mining* didalam toko sepatu, *data mining* ini digunakan untuk mengetahui frekuensi penjualan sepatu dan penyiapan stok sepatu dagangan yang diperlukan di kemudian hari.

Esha, A., et al (2020), Implementasi Algoritma Apriori untuk Rekomendasi Produk pada Toko Online, penelitian ini menjelaskan bahwa algoritma apriori selain dapat digunakan untuk mempersiapkan stok barang dagangan yang akan dibeli, algoritma ini juga dapat menjadi system rekomendasi untuk pembeli.

3. **© Hak cipta milik IBI KKG (Institut Bisnis dan Informatika Kwik Kian Gie)**

Hak Cipta Dilindungi Undang-Undang

Institut Bisnis dan Informatika Kwik Kian Gie



1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik dan tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IBIKKG.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IBIKKG.



BAB III

Analisis Sistem Yang Berjalan

© Hak cipta milik IBI KKG (Institut Bisnis dan Informatika Kwik Kian Gie)

Hak Cipta Dilindungi Undang-Undang

Institut Bisnis dan Informatika Kwik Kian Gie

A. Gambaran Umum Objek Penelitian

Sebuah wujud keberhasilan tak luput jauh dari asal mula dari sejarah berdirinya toko, dari mana inspirasi, motivasi, peluang yang ada, sehingga dapat diaplikasikan serta berdirilah sebuah usaha yang dinamakan toko Mandiri Cool Jaya. Toko Mandiri Cool Jaya berdiri pada tahun 1997, toko Mandiri Cool Jaya didirikan oleh bapak Sonny Lumintang,. Toko Mandiri Cool Jaya didirikan di Sunter Hijau Jakarta.

Dengan kemajuan teknologi dan zaman dari tahun ke tahun *SparePart* mulai dibutuhkan oleh banyak orang karena banyaknya orang yang memakai teknologi, maka salah satu peluang yang paling gencar pada saat ini adalah menjual *SparePart* dari teknologi dan tidak sedikit yang banyak pedagang yang mulai mengambil peluang tersebut untuk dijadikan usaha bisnis keluarga.

Asal mula berdirinya Toko Mandiri Cool Jaya berawal karena pada saat itu bapak Sonny pernah berkerja di pabrik perusahaan Sharp dan setelah menikah dia diajak temannya untuk berjualan *SparePart* untuk membantu menambah penghasilan serta menafkahi keluarganya. Setelah berhenti dari perusahaan Sharp bapak Sonny mulai memfokuskan dirinya kepada toko *SparePart* nya sendiri.

Asal mula nama Mandiri Cool Jaya untuk dijadikan nama toko tersebut adalah berasal dari istilah Mandiri yang berarti sikap untuk tidak menggantungkan keputusan kepada orang lain. Seorang yang menjalankan wirausaha harus mampu hidup mandiri tidak bergantung dengan orang lain, mampu memberikan keputusan terhadap suatu masalah dalam usahanya, lalu Cool yang melambangkan jenis *SparePart* yang dijualnya yaitu untuk pendingin, dan yang terakhir Jaya yang berarti selalu berhasil/sukses/hebat; sehingga termotivasi pemilik untuk diambil menjadi sebuah nama toko atau tempat berdirinya usaha Mandiri Cool Jaya. Dengan tujuan agar usahanya yang mandiri tersebut menjadi jaya.

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik dan tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar IBIKKG.

2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IBIKKG.



Toko *SparePart* Mandiri Cool Jaya adalah salah satu toko pendistributor *SparePart*

AC dan Kulkas di Indonesia. Toko ini telah berdiri dari tahun 1990, dan telah menjual ratusan ribu buah *SparePart*. Pelanggan toko ini datang dari pendistributor lain, perakit AC maupun kulkas, tukang servis/perawatan, dan lain lain.

Toko ini terletak di kawasan Jakarta bagian utara tepatnya di Kecamatan Tanjung Priok, Jalan Sunter Hijau. Visi dari toko ini adalah Melayani dan menjawab kebutuhan konsumen dengan empati, dengan mengedepankan kualitas dan kenyamanan konsumen. Misi dari toko ini adalah menyediakan barang dagangan yang berkualitas, memberikan pelayanan yang baik, serta menawarkan harga yang terjangkau.

B. Analisis Sistem yang Berjalan

Toko Mandiri Cool Jaya adalah sebuah toko yang bergerak dibidang penjualan *SparePart*. Toko ini menjual macam-macam *SparePart* untuk AC dan kulkas. Proses penjualan di toko masih menggunakan sistem yang manual, yaitu dengan cara konsumen/pembeli datang untuk memilih produk yang diminati dan akan dilayani oleh karyawan, karyawan menerima barang dagangan yang dipilih/dipesan oleh konsumen. Karyawan menyerahkan produk ke bagian kasir untuk melakukan penghitungan total bayar. Konsumen membayar sesuai total bayar, setelah konsumen membayar bagian kasir memberikan nota pembayaran sebagai tanda bukti bahwa konsumen sudah melakukan pembelian dan membayarnya dengan lunas. Bagian kasir membuat laporan penjualan yang akan diserahkan kepada pemilik toko.

Saat ini toko *SparePart* Mandiri Cool Jaya terdapat beberapa pegawai yang memiliki tugasnya masing-masing seperti menerima pesanan, mengecek *supply* barang dagangan, mengantar barang dagangan, dan lain lain. Toko ini menggunakan bon tertulis untuk mencatat hasil pemasukan dan pengeluaran di toko tersebut, toko tersebut menarik pelanggan baru dengan cara memperkenalkan dan membagikan kartu nama mereka ke perusahaan servis maupun perusahaan penjual *SparePart* lainnya.

Hak Cipta Dilindungi Undang-Undang.

Hak cipta milik IBI KKG (Institut Bisnis dan Informatika Kwik Kian Gie)

Institut Bisnis dan Informatika Kwik Kian Gie

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik dan tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar IBIKKG.

2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IBIKKG.



C. Metodologi Penelitian

Hak Cipta milik IBI KKG (Institut Bisnis dan Informatika Kwik Kian Gie)

Hak Cipta Dilindungi Undang-Undang

Institut Bisnis dan Informatika Kwik Kian Gie

1. Teknik Pengumpulan Data

a. Observasi

Observasi merupakan salah satu metode pengumpulan data dengan cara mengamati atau meninjau secara cermat dan langsung di lokasi penelitian untuk mengetahui kondisi yang terjadi atau membuktikan kebenaran dari sebuah desain penelitian yang sedang dilakukan. Dalam penelitian ini, data yang dibutuhkan dalam observasi adalah data hasil penjualan dari toko yang diobservasi oleh peneliti, peneliti mengumpulkan data tersebut dengan studi dokumen, Teknik pengumpulan data ini dilakukan dengan mengandalkan dokumen sebagai salah satu sumber data yang digunakan untuk melengkapi penelitian. Dokumen yang digunakan adalah dokumen penjualan, pengeluaran, dan pemasukan dalam toko. Peneliti menggunakan data tersebut untuk dimasukkan kedalam program untuk mencari hasil asosiasinya.

2. Teknik Analisis Data

Penelitian ini menggunakan teknik analisis data kualitatif, yaitu:

a. Reduksi Data.

Reduksi data merupakan bagian dari analisis yang menajamkan, menggolongkan, mengarahkan, membuang yang tidak perlu, dan mengorganisasi data dengan cara sedemikian rupa hingga simpulan-simpulan akhirnya dapat ditarik dan diverifikasi. Pada penelitian ini, penulis melakukan reduksi data dikarenakan adanya beberapa hasil data transaksi yang tidak relevan pada penelitian, sehingga diperlukannya reduksi data. Reduksi data diperlukan untuk mengkaji ulang dan mengolah setiap permasalahan serta proses yang ada pada perusahaan.

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik dan tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar IBIKKG.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IBIKKG.



b. Penyajian data

Penyajian data dibatasi sebagai kumpulan informasi tersusun yang memberi kemungkinan adanya penarikan kesimpulan dan pengambilan tindakan. Dengan penyajian tersebut akan dapat dipahami apa yang harus dilakukan, menganalisis ataukah tindakan berdasarkan pemahaman yang didapat dari penyajian-penyajian tersebut. Setelah penulis mendapatkan data, selanjutnya data tersebut dikumpulkan dan disusun untuk mempermudah proses pengambilan keputusan dan penarikan kesimpulan. Data yang disajikan oleh objek penelitian dipakai oleh penulis untuk dilakukannya langkah penarikan kesimpulan.

c. Penarikan kesimpulan

Penarikan kesimpulan merupakan langkah yang penting untuk menjawab masalah yang ada dari suatu penelitian dalam bentuk hipotesis atau jawaban sementara. Kesimpulan awal yang dikemukakan masih bersifat sementara, dan akan mengalami perubahan apabila tidak ditemukan bukti-bukti yang kuat yang mendukung pada tahap pengumpulan data berikutnya. Dengan demikian kesimpulan dalam penelitian kualitatif mungkin dapat menjawab rumusan masalah yang dirumuskan sejak awal, tetapi mungkin juga tidak. Kesimpulan sementara yang dapat penulis tarik berdasarkan hasil pengumpulan data yang sudah direduksi yaitu sistem pembelian yang manual sehingga karena ada begitu banyaknya data transaksi toko tidak dapat memperhatikan barang mana saja yang paling banyak dibeli dan tidak dapat mengetahui preferensi pembeli dengan cepat.

3. Teknik Pengukuran Data

Teknik pengukuran data yang dilakukan oleh penulis adalah hasil pengumpulan dan analisis data yang dikaji, dapat digunakan dalam penelitian ini untuk menilai

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik dan tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar IBIKKG.

2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IBIKKG.

apakah data yang diberikan merupakan data yang valid. Pengumpulan data dengan metode kualitatif memberikan data yang harus direduksi terlebih dahulu untuk mengurangi data yang tidak relevan dengan penelitian. Data yang sudah melewati dua proses sebelumnya dapat ditarik kesimpulannya, maka data yang diperoleh adalah data yang valid.

D. Jadwal Penelitian

No.	Jenis Kegiatan	week							
		1	2	3	4	5	6	7	8
1	pengajuan Judul								
2	penyusunan proposal								
3	pengumpulan data								
4	revisi proposal								
5	penyusunan instrumen								
6	analisis data								
7	penyusunan skripsi								

© Hak cipta milik IBI KKG (Institut Bisnis dan Informatika Kwik Kian Gie)

Institut Bisnis dan Informatika Kwik Kian Gie

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik dan tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IBIKKG.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IBIKKG.



Daftar Pustaka

- Richard J.R. (2017), *Data Mining A Tutorial-Based Primer SECOND EDITION*, Minnesota: CRC Press.
- David, D.O. (2017), *Descriptive Data Mining*, USA: Springer.
- Patricia L. (2017), *Research Design: Quantitative, Qualitative, Mixed Methods, Arts-Based, and Community-Based Participatory Research Approaches*, NY: The Guilford Press.
- Paateek Bhatia (2019), *Data Mining and Data Warehousing Principles and Practical Techniques*, India: Cambridge University Press.
- Galit, S et al (2018), *DATA MINING FOR BUSINESS ANALYTICS Concepts, Techniques, and Applications in R*, USA: John Wiley & Sons, Inc.
- Francesco Corea (2019), *An Introduction to Data Everything You Need to Know About AI, Big Data and Data Science*, Switzerland: Springer.
- Esha Alma'arif. et al (2020), "Implementasi Algoritma Apriori untuk Rekomendasi Produk pada Toko Online", *Citec Journal*, Vol. 7, No.1.
- Erma Delima Sikumbang (2018), "Penerapan Data Mining Penjualan Sepatu Menggunakan Metode Algoritma Apriori", *Jurnal Teknik Komputer* Vol 4, No. 1.
- J. Chandira dan K.R. Dewi (2019) "Implementation of Data Mining Sales of Milk Using Apriori Algorithm Method", IOP Publishing.
- O. Chouat dan H. Irawan (2018) "Implementation of Data Mining on Online Shop in Indonesia", IOP Publishing.

