

BAB II

TINJAUAN PUSTAKA

2.1 Doom Spending

Doom spending adalah istilah yang merujuk pada perilaku belanja impulsif yang dilakukan oleh individu tanpa mempertimbangkan konsekuensi jangka panjang terhadap kesehatan finansial mereka. Istilah ini sering kali dihubungkan dengan pengeluaran yang berlebihan dan tidak terencana, dimana individu merasa tertekan untuk memenuhi ekspektasi sosial atau mengikuti tren yang sedang populer. Dalam konteks ini, *doom spending* dapat mencakup :

1. Pengeluaran emosional belanja yang dilakukan sebagai respon terhadap stress, kecemasan, atau perasaan negatif lainnya, yang sering kali tidak memperhatikan anggaran atau kebutuhan finansial.
2. Pengaruh media sosial dorongan untuk membeli barang-barang tertentu sebagai upaya untuk mendapatkan pengakuan atau validasi di platform media sosial.
3. Tekanan sosial kecenderungan untuk membeli barang-barang yang dianggap “harus dimiliki” agar sesuai dengan norma atau harapan kelompok sebaya.
4. Akumulasi utang dampak jangka panjang dari *doom spending* dapat mencakup akumulasi utang yang signifikan dan penurunan kualitas hidup dapat kesulitan finansial.

Secara keseluruhan, *doom spending* mencerminkan perilaku konsumsi yang dapat merugikan individu, terutama di era digital dimana iklan dan pengaruh sosial sangat kuat [4].

2.2 Data Mining

2.2.1 Pengertian Data Mining

Data mining adalah sebagai proses untuk mendapatkan informasi yang berguna dari gudang basis data yang besar. Data mining juga dapat diartikan sebagai pengekstrakan informasi baru yang diambil dari bongkahan data besar yang membantu dalam pengambilan keputusan. Istilah data mining

kadang disebut juga Knowledge Discovery Salah satu teknik yang dibuat dalam data mining adalah bagaimana menelusuri data yang ada untuk membangun sebuah model, kemudian menggunakan model tersebut agar dapat mengenali pola data yang lain yang tidak berada dalam basis data yang tersimpan. Kebutuhan untuk prediksi juga dapat memanfaatkan teknik ini. Dalam data mining, pengelompokan data juga bisa dilakukan [5].

2.2.2 Fungsi Data Mining

1. Prediction

Fungsi data mining yang pertama adalah prediksi atau *prediction*. Ini adalah proses untuk menemukan pola dari data dengan dan juga menggunakan beberapa variabel untuk memprediksikan variabel lainnya yang nilai atau jenisnya masih tidak diketahui[6].

2. Description

Berikutnya ada fungsi deskripsi atau *description*. Ini adalah proses untuk menemukan suatu ciri krusial dari data yang terdapat di dalam suatu *database* atau basis data[7].

3. Classification

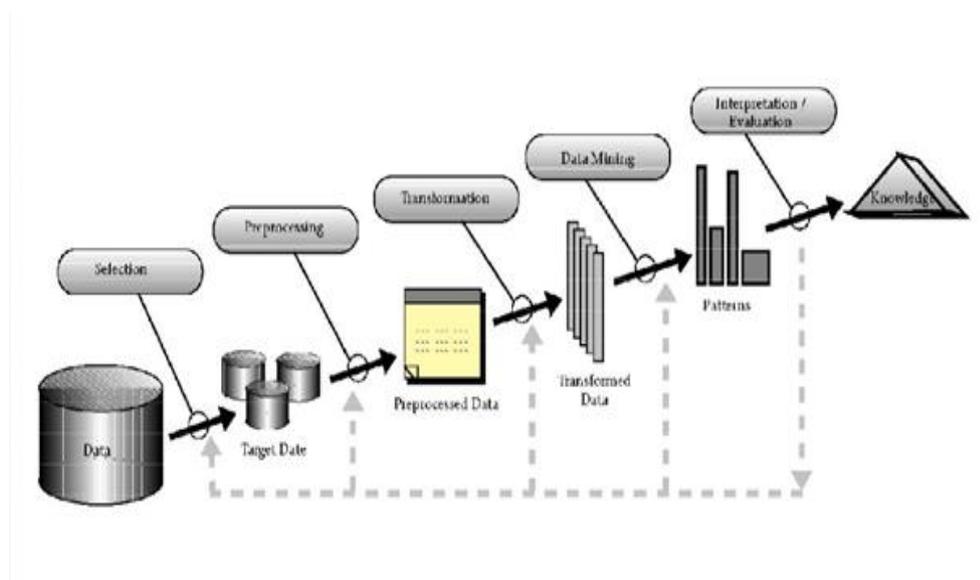
Klasifikasi atau *classification* adalah suatu proses untuk menemukan contoh atau fungsi agar dapat menggambarkan grup atau konsep dari suatu data. Proses yang digunakan untuk menggambarkan data tersebut adalah hal yang penting serta juga bisa memprediksi kecenderungan data yang terdapat pada masa depan (mendatang)[8].

4. Association

Yang terakhir adalah asosiasi atau *association*. Ini adalah proses yang dipakai untuk menemukan suatu hubungan yang terdapat pada nilai atribut daripada sekumpulan data[9].

2.2.3 Tahapan Dalam Data Mining

Tahapan yang dilakukan pada proses data mining diawali dari seleksi data dari data sumber ke data target, tahap *preprocessing* untuk memperbaiki kualitas data, transformasi, data mining serta tahap interpretasi dan evaluasi yang menghasilkan output berupa pengetahuan baru yang diharapkan memberikan kontribusi yang lebih baik [6]. Secara detail dijelaskan sebagai berikut:



Gambar 2.1 Tahap-tahap Data Mining

Tahap-tahap data mining adalah sebagai berikut:

1. Data selection

Pemilihan (seleksi) data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam KDD dimulai. Data hasil seleksi yang digunakan untuk proses data mining, disimpan dalam suatu berkas, terpisah dari basis data operasional.

2. Pre-processing /cleaning

Sebelum proses data mining dapat dilaksanakan, perlu dilakukan proses *cleaning* pada data yang menjadi fokus KDD. Proses *cleaning*

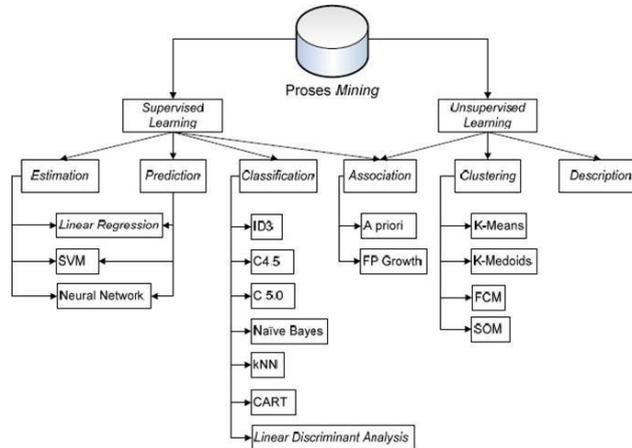
mencakup antara lain membuang duplikasi data, memeriksa data yang konsisten, dan memperbaiki kesalahan pada data.

3. Transformation

Proses transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses data mining. Proses coding dalam KDD merupakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam basis data.

4. Data mining

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.



Gambar 2.2 Beberapa metode Data Mining

5. Interpretation/evaluation

Pola informasi yang dihasilkan dari proses *data mining* perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap ini merupakan bagian dari proses KDD yang disebut *interpretation*.

2.3 Clustering

Clustering/Klasterisasi merupakan teknik mengelompokkan data pada *data base* yang mengolah banyak data dalam database tersebut yang berdasarkan kriteria yang telah ditentukan sebelumnya. Pengelompokan data pada metode *clustering* menentukan *cluster* tanpa berdasarkan kelas- kelas tertentu. *Cluster* juga dapat dipakai untuk mengelompokkan data yang kelasnya belum diketahui sama sekali. Bahkan *clustering* dapat digunakan untuk memberikan label pada kelas data yang belum diketahui, sehingga *clustering* dapat digolongkan kedalam metode *unsupervised learning*. Konsep dari *clustering* sangatlah sederhana yaitu mengelompokkan objek kedalam *cluster* yang memiliki kesamaan, semakin tinggi kesamaan dari objek tersebut maka semakin akurat juga hasil dari *cluster* tersebut[5].

Clustering biasanya digunakan dalam intelijen bisnis, pengenalan pola visual, pencarian *web*, *biosains*, dan keamanan. *Clustering* adalah proses pengelompokan data menjadi beberapa *cluster* sehingga data dalam suatu *cluster* memiliki kemiripan yang maksimal. Cara kerja teknik ini adalah dengan mengelompokkan kumpulan data ke dalam kelas-kelas atau *cluster*, dimana objek-objek pada kelas tersebut mempunyai kemiripan yang tinggi dengan objek-objek lain pada kelas tersebut, namun memiliki kemiripan yang rendah dengan objek-objek lain pada kelas tersebut di kelas/*cluster*lain [10].

2.4 K-Means Clustering

K-Means clustering merupakan salah satu metode data *clustering* non-hirarki yang mengelompokkan data dalam bentuk satu atau lebih *cluster*/kelompok. Data-data yang memiliki karakteristik yang sama dikelompokkan dalam satu *cluster*/kelompok dan data yang memiliki karakteristik yang berbeda dikelompokkan dengan *cluster*/kelompok yanglain sehingga data yang berada dalam satu *cluster*/kelompok memiliki tingkat variasi yang kecil [11].

Tahapan untuk perhitungan algoritma K-Means dijelaskan di bawah ini:

1. Tetapkan banyaknya jumlah cluster (k)
2. Pilih random titik pusat untuk cluster (centroid)
3. Pakai rumus Euclidean Distance untuk mendapatkan jarak tiap data terhadap centroid, rumusnya yakni:

$$(x,y) = \sqrt{\sum_{n=1}^i (x_i - y_i)^2}$$

Keterangan:

(x,y): jarak data x dan y

x: titik data objek

y: titik data centroid

i: banyaknya objek

4. Kelompokkan data yang sudah dikalkulasikan menurut jarak terkecil (minimum) antara data tersebut dengan pusat cluster atau data centroid dan mendapatkan cluster baru.
5. Lakukan perhitungan kembali berdasarkan data yang mengikuti cluster masing-masing untuk pusat cluster (centroid) baru. Nilai centroid baru didapatkan dari hasil perhitungan rata-rata data terhadap setiap cluster.
Rumus sebagai Berikut:

$$C1 = \frac{x_1 + x_2 + x_3 + \dots + x_n}{\sum x}$$

Keterangan:

CI : centroid baru

x1 : nilai cluster ke-1

xn : nilai cluster ke-n

x : jumlah data

6. Setelah dapat centroid baru, maka lakukan iterasi selanjutnya atau ulangi langkah c samapai e sampai tidak ditemukan data yang berpindah-pindah dari cluster [11].

2.5 Google Colaboratory

Colaboratory, atau “Colab” merupakan produk dari Google Research. Colab memungkinkan siapa saja menulis dan mengeksekusi kode python arbitrer melalui browser, dan sangat cocok untuk machine learning, analisis data, serta pendidikan. Secara lebih teknis, Colab merupakan layanan notebook Jupyter yang dihosting dan dapat digunakan tanpa penyiapan, serta menyediakan akses gratis ke resource komputasi termasuk GPU. Resource Colab tidak dijamin dan sifatnya terbatas, serta batas penggunaannya terkadang berfluktuasi. Hal ini diperlukan agar Colab dapat menyediakan resource secara gratis.

Pengguna yang ingin memiliki akses lebih andal ke resource yang lebih baik dapat menggunakan Colab Pro. Memperkenalkan Colab Pro merupakan langkah pertama yang Google ambil untuk melayani pengguna yang ingin melakukan lebih banyak hal di Colab. Tujuan jangka panjang pihak Google adalah untuk terus menyediakan versi gratis Colab, dan di saat yang bersamaan berkembang secara berkelanjutan untuk memenuhi kebutuhan pengguna Google [12].

2.6 Google Looker Studio

Google Looker Studio (sebelumnya dikenal sebagai Google Data Studio) adalah platform visualisasi data dan pelaporan interaktif yang memungkinkan pengguna untuk membuat laporan dan dasbor (dashboard) yang mudah dipahami dan dapat disesuaikan. Looker Studio memungkinkan pengguna untuk menghubungkan berbagai sumber data, termasuk Google Analytics, Google Sheets, BigQuery, dan sumber data lainnya, untuk menghasilkan visualisasi data yang dinamis dan interaktif. Pengguna dapat membuat grafik, tabel, peta, dan banyak elemen visual lainnya [13].

2.7 Penelitian Terdahulu

Berikut adalah beberapa ringkasan dari penelitian sebelumnya yang terkait dengan penelitian ini :

Table 2.1 Penelitian Terdahulu

No	Judul	Metode	Tujuan	Penulis/ Tahun	Hasil
1	Analisis Pola Perilaku Konsumen dan Kebiasaan Belanja Menggunakan Metode K-Means Clustering	Metode yang Digunakan Dalam Penelitian ini adalah K-Means Clustering	Melakukan Clustering Berdasarkan Pola Perilaku Konsumen dan Kebiasaan Belanja yang Berbeda-beda.	Nazelda Raina Noperahila, Feri Sulianta/ 2024	Hasil penelitian menunjukkan bahwa data konsumen dapat dikelompokkan menjadi tiga cluster berdasarkan pola perilaku belanja mereka. Cluster pertama terdiri dari konsumen yang sering berbelanja setiap bulan dan membeli blouse. Cluster kedua adalah konsumen yang berbelanja setiap minggu, sedangkan cluster ketiga adalah konsumen yang berbelanja setiap tiga bulan sekali.
2	Analisa Data Shopping Trends Menggunakan Algoritma	algoritma klasifikasi Naive Bayes	menganalisis tren belanja menggunakan algoritma klasifikasi	Said Imam Puro, Andi Diah Kuswanto, Jodi Hariyan,	Hasil penelitian menunjukkan bahwa metode Naive Bayes efektif dalam menemukan pola dan

	Klasifikasi Dengan Metode Naive Bayes		Naive Bayes untuk memahami perilaku konsumen di sektor ritel.	Ridho Rafliansyah, Muhammad Rival Aziz, dan Pebro Vaulina Rajagukguk/ 2024.	tren belanja konsumen berdasarkan atribut- atribut yang relevan.
--	--	--	--	--	---