

# Penerapan Algoritma C45 untuk Penilaian Karyawan pada Restoran Cepat Saji

Harry Dhika\*, Fitriana Destiawati

Jurusan Informatika, FTMIPA

Universitas Indraprasta PGRI

Jl. Nangka No. 58A, Jakarta

dhikatr@yahoo.com

honeyzone86@gmail.com

**Abstrak**—Kajian dilakukan pada restoran cepat saji untuk mengetahui atau memberikan nilai pada karyawannya sehingga diperoleh kesimpulan bahwa karyawan tersebut memberikan *point* yang baik, sangat baik dan memprediksi kurang baik. Karyawan merupakan salah satu aset yang dimiliki oleh perusahaan dalam hal ini restoran cepat saji, melihat dan menjaga kualitas dari karyawan tentunya merupakan salah satu faktor penentu keberlangsungan restoran cepat saji, untuk itu pentingnya dilakukan kajian terhadap karyawan guna menjaga dan meningkatkan kualitas restoran cepat saji. Memberikan *reward* kepada karyawan dengan mengkaji terlebih dahulu merupakan apresiasi yang pantas diterima bagi aset perusahaan yakni karyawan. Metode kajian menggunakan konsep *data mining* menggunakan algoritma C45 yang digunakan dalam melakukan penerapan data karyawan sehingga diketahui struktur pohon penilaian karyawan hingga memprediksi pola karyawan yang ada. Terdapat data yang digunakan untuk mengolah data dan melakukan evaluasi data karyawan. Hasil dari pengolahan dievaluasi dengan konsep *confusion matrix*. Dengan metode algoritma C45, tingkat akurasi diperoleh sangat tinggi yakni 97,71%. Kajian dilakukan pada restoran cepat saji di kota Bogor dengan melihat jumlah karyawan yang besar sesuai dengan komposisi *data mining*.

**Kata kunci**—Algoritma C45, *data mining*, restoran cepat saji, karyawan

## I. PENDAHULUAN

Restoran cepat saji merupakan objek kajian, khususnya karyawan di dalamnya. Hal ini dikarenakan pentingnya menjaga karyawan pada restoran cepat saji salah satunya dengan tujuan untuk menjaga kualitas penyajian dan pelayanan demi keberlangsungannya dalam bisnis perusahaan. Pekerjaan yang paling padat di mana hampir tidak bisa duduk atau beristirahat sejenak terjadi pada restoran cepat saji, dengan begitu tingkat kepadatan, tingkat kejenuhan sangat tinggi. Atas dasar ini maka perlunya menjaga salah satu aset dari perusahaan di luar dari aset tetap seperti bangunan dan tanah yakni karyawan. Pentingnya penghargaan pada karyawan restoran cepat saji merupakan salah satu apresiasi perusahaan untuk menjaga keberlangsungan usahanya.

Kajian terdahulu yang relevan dan sejenis terhadap restoran cepat saji yakni kajian mengenai kepuasan pelanggan pada restoran cepat saji yang lebih fokus membahas mengenai kepuasan konsumen dalam santap pada menu restoran cepat saji[1]. Kajian lain yang sejenis yakni kajian kepuasan

konsumen terhadap restoran cepat saji dengan pendekatan *data mining*, menggunakan metode[2]. Kajian yang dilakukan dalam naskah ini Decision Tree lebih pada sudut pandang yang berbeda dari segi penentuan karyawan terbaik, sebagai apresiasi terhadap karyawan dalam melakukan tugasnya.

Tujuan kajian yakni untuk menerapkan algoritma C45 dalam data restoran cepat saji, dengan demikian dilakukan pengembangan untuk terus meningkatkan akurasi, agar dapat di prediksi sesuai hasil akhir bahwa penerapan algoritma C45 dengan konsep *data mining* dapat menentukan dengan akurat prediksi karyawan baik, kurang baik dan sangat baik.

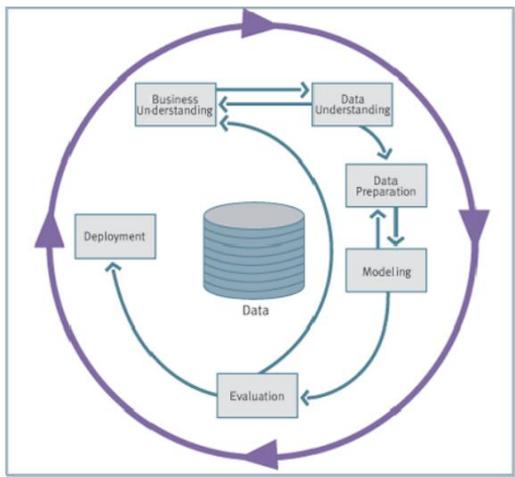
## II. METODE

### A. Data Mining

Kajian ini menggunakan konsep *data mining*, *data mining* atau menggali data diartikan sebagai metode yang pakai untuk mengekstraksi informasi prediktif tersembunyi pada *database*, ini adalah pengetahuan yang sangat berguna bagi perusahaan dalam memberdayakan data *warehouse*[3]. *Data mining* disebut sebagai proses ekstraksi pengetahuan dari data yang besar, sesuai fungsinya *data mining* adalah proses pengambilan keputusan dari *volume* data yang besar yang disimpan dalam basis data, data *warehouse*, atau informasi yang disimpan dalam *repository*[4].

Istilah *data mining* berasal dari kemiripan antara pencarian informasi yang bernilai dari basis data yang besar dengan menambang sebuah gunung untuk memperoleh sesuatu yang bernilai[5]. Nilai baru nantinya akan dapat digunakan kembali dan diimplementasikan pada lingkungan baru yang sejenis, dengan begitu terbentuk *rule* dari pola yang sudah ada pada restoran cepat saji. Pola baru ini merupakan *habit* atau kebiasaan dalam perusahaan tertentu dalam melakukan kajian terhadap pegawainya khususnya penilaian karyawan terbaik

Kajian ini menggunakan desain dengan konsep CRISP-DM (*Cross Standart Industries for Data Mining*), yakni terdiri dari *Business/Research Understanding Phase*, *Data Understanding Phase* (Fase Pemahaman Data), *Data Preparation Phase* (Fase Pengolahan Data), *Modeling Phase* (Fase Pemodelan), *Evaluation Phase* (Fase Evaluasi), *Deployment Phase* (Fase Penyebaran) dalam metode ini terdapat 6 tahapan[6][7]:



Gambar 1. Tahap CRISP-DM (Cross Standart Industries for Data Mining)

Dari tahapan tersebut di atas jika diterjemahkan maka *Business/ Research Understanding Phase* Dalam hal ini dilakukan pemahaman terhadap penelitian yang dilakukan perlunya pemahaman terhadap substansi atau inti dari penelitian yang dilakukan mulai dari kebutuhan dan perspektif bisnis yang dilakukan, terdapat beberapa kegiatan pada tahapan ini diantaranya adalah ditentukannya sasaran dan tujuan dari penelitian, pemahaman kondisi atau situasi bisnis, menentukan tujuan dari *data mining* dan melakukan penjadwalan atau perencanaan strategi penelitian.

*Data Understanding Phase* (Fase Pemahaman Data). Dikenal dengan fase pemahaman terhadap data yang diperoleh kemudian data awal yang dikumpulkan perlu dilakukan kajian sehingga diketahui data yang akan digunakan. Dalam tahapan ini perlunya identifikasi terhadap masalah kualitas data yang baik sehingga diperoleh nilai subset yang sesuai, menarik dalam pembuatan hipotesa awal.

*Data Preparation Phase* (Fase Pengolahan Data). Tahapan ini merupakan tahapan pengolahan data atau dapat juga dikatakan sebagai tahapan persiapan data. Banyak persiapan yang dilakukan pada tahapan ini tak jarang fase ini juga disebut sebagai fase padat karya. Beberapa kegiatan seperti pemilihan tabel dan *field* terjadi pada fase ini. Pemilihan tabel dan *field* tersebut akan di masukan atau di transformasikan ke dalam *database* yang lain atau *database* baru sebagai bahan atau data mentah, *data mining* mentah.

*Modeling Phase* (Fase Pemodelan). Pada fase pemodelan dilakukan penggunaan aplikasi seperti *Rapidminer*, aplikasi pengolah *data mining* mentah, dimasukkan juga algoritma C4.5, dari data tersebut dipilih atribut yang menjadi label, seluruh parameter dipilih dengan penentuan nilai yang optimal.

*Evaluation Phase* (Fase Evaluasi). Fase ini merupakan tahapan analisa dari hasil pengolahan fase sebelumnya dengan menginterpretasikan data yang kemudian diperoleh nilai perbandingan dengan proses model yang sebelumnya. Perlunya kajian mendalam pada tahapan ini untuk menentukan nilai akurasi data model yang dihasilkan hal tersebut bertujuan agar dapat digunakan oleh sasaran sesuai rencana pada *domain goal* pada fase pertama

*Deployment Phase* (Fase Penyebaran). Fase ini merupakan tahapan pembuatan laporan atau implementasi *knowledge* yang diperoleh pada fase sebelumnya.

### B. Algoritma C45

Algoritma C4.5 merupakan algoritma pengklasifikasi dengan teknik pohon keputusan yang terkenal dan disukai karena memiliki kelebihan-kelebihan. Kelebihan ini misalnya dapat mengolah data numerik (kontinu) dan diskrit, dapat menangani nilai atribut yang hilang, menghasilkan aturan-aturan yang mudah diinterpretasikan dan tercepat diantara algoritma-algoritma yang lain. Keakuratan prediksi yaitu kemampuan model untuk dapat memprediksi label kelas terhadap data baru atau yang belum diketahui sebelumnya dengan baik. Dalam hal kecepatan atau efisiensi waktu komputasi yang diperlukan untuk membuat dan menggunakan model. Kemampuan model untuk memprediksi dengan benar walaupun data ada nilai dari atribut yang hilang, dan juga skalabilitas yaitu kemampuan untuk membangun model secara efisien untuk data berjumlah besar (aspek ini akan mendapatkan penekanan). Terakhir interpretabilitas yaitu model yang dihasilkan mudah dipahami. Selain itu dapat juga dikatakan bahwa Algoritma merupakan kumpulan perintah yang tertulis secara sistematis guna menyelesaikan permasalahan logika dari matematika. Pengertian algoritma C4.5 merupakan algoritma yang digunakan untuk membentuk pohon keputusan. Sedang pohon keputusan dapat diartikan suatu cara untuk memprediksi atau mengklarifikasi yang sangat kuat. Pohon keputusan dapat membagi kumpulan data yang besar menjadi himpunan-himpunan *record* yang lebih kecil dengan menerapkan serangkaian aturan keputusan

Ada beberapa tahap dalam membuat sebuah pohon keputusan dengan algoritma C4.5[8]:

1. Menyiapkan data *training*. Data *training* biasanya diambil dari data histori yang pernah terjadi sebelumnya dan sudah dikelompokkan ke dalam kelas-kelas tertentu.
2. Menentukan akar dari pohon. Akar akan diambil dari atribut yang terpilih, dengan cara menghitung nilai *gain* dari masing-masing atribut, nilai *gain* yang paling tinggi yang akan menjadi akar pertama. Sebelum menghitung nilai *gain* dari atribut, hitung dahulu nilai *entropy*. Untuk menghitung nilai *entropy* digunakan persamaan(1):

$$Entropy(S) = \sum_{i=1}^n - p_i \cdot \log_2 p_i \quad (1)$$

Keterangan:

S = himpunan kasus

n = jumlah partisi S

$p_i$  = proporsi  $S_i$  terhadap S

3. Kemudian hitung nilai *gain* menggunakan persamaan(2):

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \quad (2)$$

Keterangan:

- S = himpunan kasus
- A = fitur
- n = jumlah partisi atribut A
- $|S_i|$  = proporsi  $S_i$  terhadap S
- $|S|$  = jumlah kasus dalam S

4. Ulangi langkah ke-2 hingga semua *record* terpartisi.
5. Proses partisi pohon keputusan akan berhenti saat :
  - a. Semua *record* dalam simpul N mendapat kelas yang sama.
  - b. Tidak ada atribut di dalam *record* yang dipartisi lagi.
  - c. Tidak ada *record* di dalam cabang yang kosong.

Evaluasi menggunakan *confusion matrix* dengan rumusan pada Tabel 1[9]:

TABEL 1. TABEL CONFUSION MATRIX

Klasifikasi yang benar	Diklasifikasikan sebagai	
	+	-
+	true positives	false negatives
-	false positives	true negatives

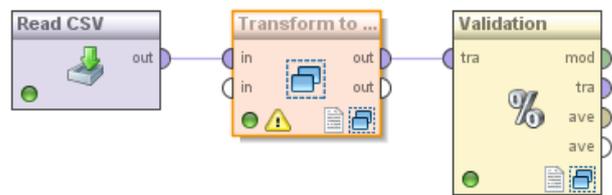
### III. HASIL DAN DISKUSI

Kajian dilakukan dengan menggunakan perangkat lunak Rapidminer 5.1. dengan menampilkan atribut pada Tabel 2:

TABEL 2. ATRIBUT DATA

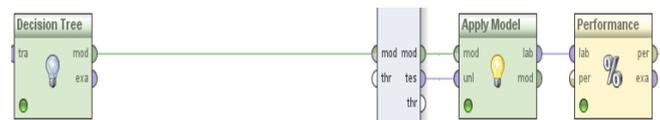
Role	Nama Atribut	Tipe
Label	Jam Kerja	Polynomial
Regular	Kreatifitas	Polynomial
Regular	Disiplin	Polynomial
Regular	Bermasalah	Polynomial
Regular	Prestasi	Polynomial
Regular	Etika	Polynomial
Regular	Tanggung Jawab	Binomial
Regular	Kepemimpinan	Polynomial
Regular	Inisiatif	Polynomial
Regular	Kerjasama	Polynomial
Regular	Kebersihan	Binomial

Dari *rule* atribut tersebut dilakukan pengolahan data sebanyak 437 record pada Rapidminer dengan pola sebagai pada Gambar 2.



Gambar 2. Pola Pengolahan dengan Rapidminer (Main Proses)

Gambar 1. Menunjukkan proses pembacaan terhadap data dengan format CSV kemudian mengubah data menjadi nominal, polynomial atau binomial dan melakukan validasi untuk mengukur nilai akurasinya. Penerapan algoritma C45 akan dilakukan pada proses validasi, proses akurasi juga dilakukan dalam tahapan validasi yang lebih spesifik lagi seperti pada Gambar 2:



Gambar 3. Proses dengan Algoritma C45 dan Pengukuran Tingkat Akurasi

Seluruh proses dilakukan dengan tahapan CRIP-DM sehingga masukan dengan algoritma C45 kemudian dilakukan evaluasi menghasilkan memberikan hasil seperti pada Gambar 4.

accuracy: 97.71% +/- 1.45% (mikroc: 97.71%)				
	true Baik	true Kurang Baik	true Sangat Baik	class precision
pred. Baik	239	6	0	97.55%
pred. Kurang Baik	4	53	0	92.98%
pred. Sangat Baik	0	0	135	100.00%
class recall	98.35%	89.83%	100.00%	

Gambar 4. Hasil Akurasi

Hasil pengukuran akurasi data diperoleh bahwa secara keseluruhan nilainya mencapai 97,71% dari tabel tersebut diketahui prediksi Baik, karyawan dengan true Baik mencapai nilai 239 karyawan dan untuk true Kurang Baik sebanyak 6 Karyawan dan true Sangat Baik sebanyak 0 Karyawan dengan hasil capaian nilai presisi mencapai sebesar 97,55% sedangkan untuk Prediksi Kurang Baik data karyawan untuk true Baik mencapai nilai 4 konsumen, untuk true Kurang Baik terdapat 53 Karyawan dan True Sangat Baik Sebanyak 0 Karyawan, persentase prediksi kurang baik adalah sebesar 92,98% Karyawan untuk class Presisinya. Pada Prediksi Sangat Baik Nilai True Baiknya jumlahnya 0 karyawan, jumlah true kurang baiknya adalah 0 karyawan, dan prediksi sangat baik untuk true sangat baik mencapai 135 karyawan, dan ini merupakan jumlah karyawan teladan dengan class presisi mencapai sempurna 100%. Untuk class recall data karyawan sangat tinggi yakni untuk class recall true baik mencapai 98,35%, untuk class recall pada jumlah karyawan true kurang baik persentasenya mencapai 89,83%, sedangkan untuk class recall true sangat baik mencapai nilai sempurna yakni 100% karyawan terbaik. Pengukuran akurasi melihat data lebih besar kepada karyawan sangat baik dengan tingkat akurasi tertinggi yang berbanding lurus dengan pengukuran *presicion*.

Dari hasil pada Gambar 3 maka dilakukan pengukuran performance Vektor sebagai berikut:

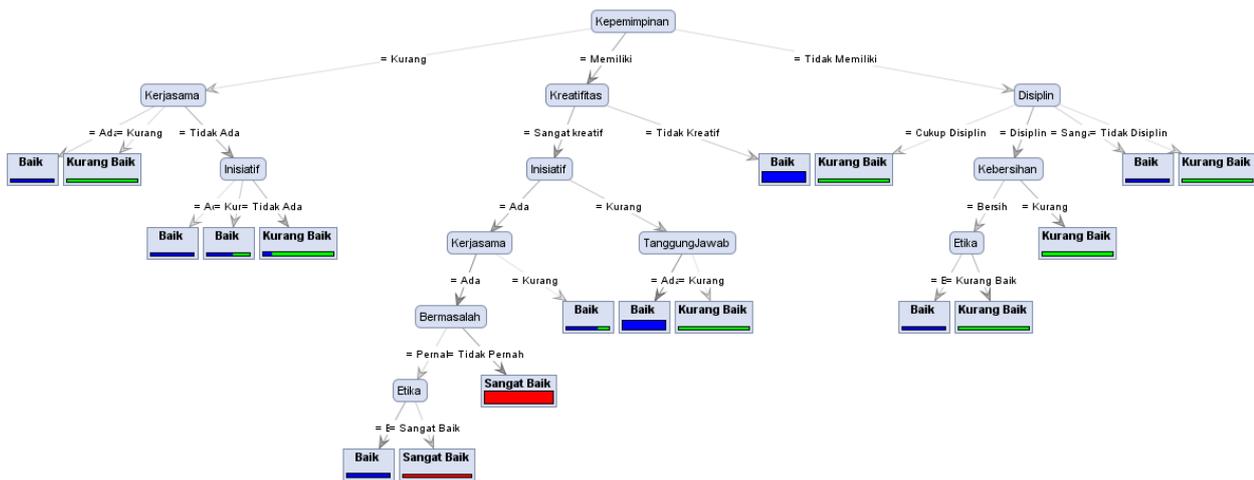
Kurang Baik:	4	53	0
Sangat Baik:	0	0	135

PerformanceVector:  
accuracy: 97.71% +/- 1.45% (mikro: 97.71%)

ConfusionMatrix:  
True: Baik Kurang Baik Sangat Baik  
Baik: 239 6 0  
Kurang Baik: 4 53 0  
Sangat Baik: 0 0 135  
kappa: 0.960 +/- 0.026 (mikro: 0.960)

Berdasarkan algoritma C45 juga dapat memberikan gambaran pohon keputusan yang disusun berdasarkan nilai *entropy* dan nilai *Gain*, memberikan klasifikasi penggolongan berdasarkan pohon atau simpul, cabang hingga membuat *leaf*, pada Gambar 5:

ConfusionMatrix:  
True: Baik Kurang Baik Sangat Baik  
Baik: 239 6 0



Gambar 5. Pohon Penilaian Karyawan Pada Restoran Cepat Saji

Dari Gambar 4 dapat disederhanakan dalam bentuk teks yang dapat lebih dipahami dengan baik, dengan menjabarkan sebagai berikut:

```

Kepemimpinan = Kurang
| Kerjasama = Ada: Baik {Baik=10, Kurang Baik=0, Sangat Baik=0}
| Kerjasama = Kurang: Kurang Baik {Baik=0, Kurang Baik=6, Sangat Baik=0}
| Kerjasama = Tidak Ada
| | Inisiatif = Ada: Baik {Baik=2, Kurang Baik=0, Sangat Baik=0}
| | Inisiatif = Kurang: Baik {Baik=5, Kurang Baik=3, Sangat Baik=0}
| | Inisiatif = Tidak Ada: Kurang Baik {Baik=2, Kurang Baik=12, Sangat Baik=0}
Kepemimpinan = Memiliki
| Kreativitas = Sangat kreatif
| | Inisiatif = Ada
| | | Kerjasama = Ada
| | | | Bermasalah = Pernah
| | | | Etika = Baik: Baik {Baik=18, Kurang Baik=0, Sangat Baik=0}
| | | | Etika = Sangat Baik: Sangat Baik {Baik=0, Kurang Baik=0, Sangat
Baik=8}
| | | | Bermasalah = Tidak Pernah: Sangat Baik {Baik=0, Kurang Baik=0, Sangat
Baik=127}
| | | Kerjasama = Kurang: Baik {Baik=3, Kurang Baik=1, Sangat Baik=0}
| | Inisiatif = Kurang

```

```

| | | TanggungJawab = Ada: Baik {Baik=86, Kurang Baik=0, Sangat Baik=0}
| | | TanggungJawab = Kurang: Kurang Baik {Baik=0, Kurang Baik=2, Sangat Baik=0}
| Kreatifitas = Tidak Kreatif: Baik {Baik=101, Kurang Baik=0, Sangat Baik=0}
Kepemimpinan = Tidak Memiliki
| Disiplin = Cukup Disiplin: Kurang Baik {Baik=0, Kurang Baik=4, Sangat Baik=0}
| Disiplin = Disiplin
| | Kebersihan = Bersih
| | | Etika = Baik: Baik {Baik=5, Kurang Baik=0, Sangat Baik=0}
| | | Etika = Kurang Baik: Kurang Baik {Baik=0, Kurang Baik=7, Sangat Baik=0}
| | Kebersihan = Kurang: Kurang Baik {Baik=0, Kurang Baik=20, Sangat Baik=0}
| Disiplin = Sangat Disiplin: Baik {Baik=11, Kurang Baik=0, Sangat Baik=0}
| Disiplin = Tidak Disiplin: Kurang Baik {Baik=0, Kurang Baik=4, Sangat Baik=0}

```

Sehingga dapat diketahui setiap cabang dan setiap *leaf* nya memiliki nilai dan jumlah karyawan yang baik, kurang baik dan jumlah karyawan sangat baik dari penjabaran gambar 4 maka dapat disebut sebagai *rule* baru atau model baru yang ditemukan untuk menyelesaikan model lain.

Berdasarkan Gambar 3 dan Gambar 4 maka dapat dibuat penggolongan pengukuran tingkat akurasi yang dapat dibagi menjadi beberapa kelompok[10]:

- a. 0.90-1.00 = klasifikasi sangat baik
- b. 0.80-0.90 = klasifikasi baik
- c. 0.70-0.80 = klasifikasi cukup
- d. 0.60-0.70 = klasifikasi buruk
- e. 0.50-0.60 = klasifikasi salah

Dari nilai hasil pengukuran akurasi maka dapat disimpulkan sesuai dengan pengelompokan klasifikasi masuk pada **klasifikasi sangat baik** untuk Penerapan algoritma C45 pada karyawan restoran cepat saji.

#### IV. KESIMPULAN

Penerapan algoritma C45 pada karyawan restoran cepat saji memiliki tingkat akurasi yang termasuk dalam **klasifikasi sangat baik** yakni 97,71% atau sekitar 135 karyawan dengan class presisi dan prediksi mencapai 100%, sedangkan yang baik sebanyak 239 karyawan presisinya sebesar 97%, karyawan yang kurang baik berjumlah 53 karyawan dengan presisi 92%. Kategori penelitian ini masuk dalam klasifikasi sangat baik dengan persentase 0.90-1.00.

Dengan tingkat akurasi yang sangat baik maka penilaian karyawan pada restoran cepat saji dapat dilakukan dengan *rule* yang telah dihasilkan menggunakan algoritma C45. Ke depannya *rule* tersebut dapat dijadikan model dalam pengembangan perangkat lunak untuk kepentingan restoran cepat saji.

Dari hasil tersebut maka perlu terus dikembangkan dan dikomparasikan dengan beberapa algoritma lain dan beberapa metode lain seperti SEMMA dan KDD sehingga hasil akurasi bisa jauh lebih sempurna lagi.

#### DAFTAR PUSTAKA

- [1] H. Dhika, F. Destiwati, and A. Fitriansyah, "Implementasi Algoritma C4.5 Terhadap Kepuasan Pelanggan," *SNAPP 2016*, vol. 6, no. Sains dan Teknologi, pp. 16–22, 2016.
- [2] V. Mandasari, B. A. Tama, and U. Sriwijaya, "Analisis Kepuasan Konsumen Terhadap Restoran Cepat Saji Melalui Pendekatan Data Mining : Studi Kasus XYZ," *J. Generik*, vol. 6, no. 1, pp. 4–7, 2011.
- [3] M. K. A. J. P. Jiawei Han, "Data Mining: Concepts and Techniques, Third Edition - Books24x7," *Morgan Kaufmann Publishers*, 2012. .
- [4] J. J. . Han, M. M. . Kamber, and J. Pei, *Data Mining: Concepts and Techniques*. 2012.
- [5] S. Sumathi and S. N. Sivanandam, "Data warehousing, data mining, and OLAP," *Stud. Comput. Intell.*, vol. 29, pp. 21–73, 2007.
- [6] C. McGregor, C. Catley, and A. James, "A process mining driven framework for clinical guideline improvement in critical care," in *CEUR Workshop Proceedings*, 2011, vol. 765.
- [7] A. Azevedo and M. F. Santos, "KDD, SEMMA and CRISP-DM: a parallel overview," *IADIS Eur. Conf. Data Min.*, no. January, pp. 182–185, 2008.
- [8] S. Goswami, S. Chakraborty, S. Ghosh, A. Chakrabarti, and B. Chakraborty, "A review on application of data mining techniques to combat natural disasters," *Ain Shams Eng. J.*, 2016.
- [9] M. Bramer, *Principles of Data Mining*. 2007.
- [10] F. Gorunescu, "Data mining: Concepts, models and techniques," *Intell. Syst. Ref. Libr.*, vol. 12, 2011.