

PRINSIP KLASIFIKASI DAN DATA MINING DENGAN ALGORITMA C4.5

Mamok Andri Senubekti^{*1}, Lies Anggi Puspita Dewi²

¹Program Studi Manajemen Informatika, Akademi Manajemen Informatika Dan Komputer HASS

²Program Studi Manajemen, Sekolah Tinggi Ilmu Ekonomi STEMBI
e-mail: ^{*1}mrymodion@gmail.com, ²liesanggipuspidewi@gmail.com

ABSTRAK

Pertumbuhan yang cepat dan integrasi database memberikan ilmuwan, insinyur, dan pebisnis dengan sumber daya baru yang luas yang dapat dianalisis untuk membuat penemuan ilmiah, mengoptimalkan sistem industri, dan mengungkap pola yang berharga secara finansial. Untuk melakukan proyek analisis data besar ini, peneliti dan praktisi telah mengadopsi algoritme mapan dari statistik, pembelajaran mesin, jaringan saraf, dan basis data dan juga telah mengembangkan metode baru yang ditargetkan pada masalah data mining besar. Principles of Data Mining oleh David Hand, Heikki Mannila, dan Padhraic Smyth memberikan pengenalan kepada praktisi dan siswa tentang berbagai algoritma dan metodologi di area yang menarik ini. Pada penelitian ini digunakan algoritma C4.5. Sifat interdisipliner bidang ini cocok dengan ketiga penulis ini, yang keahliannya mencakup statistik, database, dan ilmu komputer. Hasilnya adalah sebuah buku yang tidak hanya memberikan detail teknis dan prinsip-prinsip matematika yang mendasari metode data mining, tetapi juga memberikan perspektif yang berharga tentang keseluruhan perusahaan.

Kata Kunci: Data Mining, Algoritma C4.5, Database

ABSTRACT

The rapid growth and integration of database provides scientists, engineers, and entrepreneurs the vast new resources that can be analyzed to make scientific discoveries, optimize industrial systems, and uncover valuably financially patterns. To conduct this big data analysis project, researchers and practitioners have adopted well-established algorithms from statistics, machine learning, neural networks, and databases. In addition, it is also developed new methods targeted at big data mining problems. Data Mining principles by David Hand, Heikki Mannila, and Padhraic Smyth gives practitioners and students an introduction to the various algorithms and methodologies in this interesting area. This study also applied C4.5 algorithm. The interdisciplinary nature of the field suits this third author, his expertise spanning statistics, databases, and computer science. The result of the study is a book that contain not only principles that explain data mining methods, but also a valuable perspective of company as a whole.

Keywords: Data Mining, C4.5 Algorithm, Databases.

1. PENDAHULUAN

Ilmu mengekstraksi informasi yang berguna dari kumpulan data besar atau database dikenal sebagai data mining. Ini adalah disiplin baru, terletak di persimpangan statistik, pembelajaran mesin, manajemen data dan database, pengenalan pola, kecerdasan buatan, dan area lainnya. Semua ini berkaitan dengan aspek-aspek tertentu dari analisis data, sehingga mereka memiliki banyak kesamaan-tetapi masing-masing juga memiliki cita rasa yang berbeda, yang menekankan pada masalah dan jenis solusi tertentu.

Karena data mining mencakup berbagai topik dalam ilmu komputer dan statistik, tidak mungkin untuk mencakup semua materi yang berpotensi relevan dalam satu teks. Mengingat hal ini, penulis telah berfokus pada topik yang penulis yakini sebagai yang paling mendasar.

Dari sudut pandang pengajaran, artikel ini ditujukan untuk mahasiswa sarjana di tingkat senior (tahun terakhir), atau tingkat pascasarjana tahun pertama atau kedua, yang ingin belajar tentang prinsip-prinsip dasar data mining. Teks juga harus bernilai bagi peneliti dan praktisi yang

tertarik untuk mendapatkan pemahaman yang lebih baik tentang metode dan teknik data mining. Keakraban dengan konsep yang sangat dasar dalam probabilitas, kalkulus, aljabar linier, dan optimasi diasumsikan - dengan kata lain, latar belakang sarjana dalam disiplin kuantitatif apa pun seperti teknik, ilmu komputer, matematika, ekonomi, dll., harus memberikan latar belakang yang baik.

Sudah banyak buku lain tentang data mining di pasaran. Banyak yang menjadi sasaran komunitas bisnis secara langsung dan menekankan metode dan algoritma tertentu (seperti pengklasifikasi pohon keputusan) daripada prinsip umum (seperti estimasi parameter atau kompleksitas komputasi). Teks-teks ini cukup berguna dalam memberikan konteks umum dan studi kasus, tetapi memiliki keterbatasan dalam pengaturan kelas, karena prinsip-prinsip dasar yang mendasari sering hilang [1]. Ada teks lain tentang data mining yang memiliki cita rasa lebih akademis, tetapi sampai saat ini sebagian besar telah ditulis dari sudut pandang ilmu komputer, khususnya baik dari sudut pandang database, atau dari sudut pandang pembelajaran mesin.

Teks ini memiliki bias yang berbeda. Penulis telah berusaha memberikan pandangan dasar tentang data mining [2]. Daripada membahas aplikasi data mining tertentu secara panjang lebar (seperti, katakanlah, penyaringan kolaboratif, penilaian kredit, dan deteksi penipuan), penulis malah berfokus pada teori dan algoritma yang mendasari yang menyediakan "lem" untuk aplikasi semacam itu. Ini bukan untuk mengatakan bahwa penulis tidak memperhatikan aplikasi. Data mining pada dasarnya adalah disiplin terapan, dan dengan pemikiran ini penulis sering membuat referensi ke studi kasus dan aplikasi spesifik di mana teori dasar dapat (atau telah) diterapkan.

Dalam pandangan penulis, penguasaan data mining membutuhkan pemahaman tentang masalah statistik dan komputasi. Persyaratan untuk menguasai dua bidang keahlian yang berbeda ini menghadirkan tantangan yang cukup besar bagi siswa dan guru. Untuk ilmuwan

komputer tipikal, literatur statistik relatif tidak dapat ditembus: serangkaian jargon, asumsi implisit, argumen asimtotik, dan kurangnya detail tentang bagaimana konsep teoretis dan matematika sebenarnya diwujudkan dalam bentuk algoritma analisis data. Situasi ini secara efektif dibalik untuk ahli statistik: literatur ilmu komputer tentang pembelajaran mesin dan data mining penuh dengan diskusi tentang algoritma, pseudocode, efisiensi komputasi, dan sebagainya, seringkali dengan sedikit referensi ke model yang mendasari atau prosedur inferensi. Poin penting adalah bahwa kedua pendekatan tetap penting ketika berhadapan dengan kumpulan data yang besar. Pemahaman tentang tampilan "pemodelan matematika", dan tampilan "algoritma C4.5" sangat penting untuk memahami dengan benar kompleksitas data mining [3].

Penulis akan mengambil pandangan bahwa algoritme data mining yang menangani tugas-tugas ini memiliki empat komponen dasar:

1. Model atau Struktur Pola: menentukan struktur yang mendasari atau bentuk fungsional yang penulis cari dari data.
2. Fungsi Skor: menilai kualitas model yang dipasang.
3. Metode Optimasi dan Pencarian: optimasi fungsi skor dan pencarian pada model dan struktur pola yang berbeda.
4. Strategi Manajemen Data: menangani akses data secara efisien selama pencarian/optimasi.

Penulis telah membahas perbedaan antara model dan struktur pola. Di sisa bagian ini penulis akan membahas secara singkat tiga komponen lain dari algoritma data mining.

Teknik statistik saja mungkin tidak cukup untuk mengatasi beberapa masalah yang lebih menantang dalam data mining, terutama yang timbul dari kumpulan data yang sangat besar. Meskipun demikian, statistik memainkan peran yang sangat penting dalam data mining: ini adalah komponen yang diperlukan dalam setiap perusahaan data mining. Pada bagian ini penulis membahas beberapa interaksi antara statistik tradisional dan data mining.

Dengan kumpulan data yang besar (dan khususnya dengan kumpulan data yang sangat besar), penulis mungkin tidak mengetahui fakta langsung tentang data tersebut. Penggeledahan data yang sederhana bukanlah suatu pilihan. Ini berarti bahwa metode pencarian dan pemeriksaan yang canggih mungkin diperlukan untuk menerangi fitur-fitur yang akan mudah terlihat dalam kumpulan data kecil. Selain itu, seperti yang penulis komentari di atas, seringkali objek data mining adalah membuat beberapa kesimpulan di luar basis data yang tersedia. Misalnya, dalam database objek astronomi, penulis mungkin ingin membuat pernyataan bahwa "semua objek seperti ini berperilaku seperti itu", mungkin dengan probabilitas kualifikasi terlampir. Demikian pula, penulis dapat menentukan bahwa wilayah tertentu di suatu negara menunjukkan pola panggilan telepon tertentu. Sekali lagi, mungkin bukan panggilan dalam database yang ingin penulis buat pernyataannya. Sebaliknya itu mungkin akan menjadi pola panggilan masa depan yang ingin penulis prediksi.

2. METODE PENELITIAN

Penelitian ini merupakan penelitian deskriptif dengan pendekatan kualitatif.

Metode deskriptif kualitatif adalah metode penelitian yang berdasarkan pada filsafat postpositivisme digunakan untuk meneliti pada kondisi objek yang alamiah (sebagai lawannya adalah eksperimen) dimana peneliti adalah sebagai instrument kunci teknik pengumpulan data dilakukan secara trigulasi (gabungan), analisis data bersifat induktif/kualitatif, dan hasil penelitian kualitatif lebih menekankan makna daripada generalisasi [4].

Penelitian deskriptif kualitatif bertujuan untuk menggambarkan, melukiskan, menerangkan, menjelaskan dan menjawab secara lebih rinci permasalahan yang akan diteliti dengan mempelajari semaksimal mungkin seorang individu, suatu kelompok atau suatu kejadian [5].

Dalam penelitian kualitatif manusia merupakan instrumen penelitian dan hasil penulisannya berupakata-kata atau pernyataan yang sesuai dengan keadaan sebenarnya.

3. PEMBAHASAN

Data Mining tidak akan lengkap tanpa mengacu pada penggunaan historis istilah-istilah seperti "data mining". Pada tahun 1960-an, ketika komputer semakin banyak diterapkan pada masalah analisis data, tercatat bahwa jika Anda mencari cukup lama, Anda selalu dapat menemukan beberapa model yang sesuai dengan kumpulan data dengan baik. Ada dua faktor yang berkontribusi pada situasi ini: kompleksitas model dan ukuran himpunan model yang mungkin. kelas model yang dibahas oleh penulis sangat fleksibel (relatif terhadap ukuran kumpulan data yang tersedia), maka penulis mungkin akan dapat menyesuaikan data yang tersedia dengan baik. Namun, seperti yang penulis katakan di atas, tujuannya mungkin untuk menggeneralisasi di luar data yang tersedia; model yang cocok mungkin tidak ideal untuk tujuan ini. Selain itu, bahkan jika tujuannya adalah untuk mencocokkan data (misalnya, ketika penulis ingin menghasilkan ringkasan data yang paling akurat yang menggambarkan populasi yang lengkap), umumnya lebih baik melakukan ini dengan model sederhana. Untuk mengambil ekstrim, model kompleksitas setara dengan data mentah pasti akan cocok dengan sempurna, tetapi tidak akan menarik atau nilai.

3.1. Persiapan Data Untuk Data Pertambangan

Di banyak bidang ilmu komputer, seperti pengenalan pola, informasi pengambilan, pembelajaran mesin, penambangan data, dan kecerdasan Web, seseorang perlu menyiapkan data berkualitas dengan melakukan pra-pemrosesan data mentah. Dalam prakteknya, telah umumnya ditemukan bahwa pembersihan dan persiapan data memakan waktu sekitar 80% dari total upaya rekayasa data [6]. Oleh karena itu, persiapan data merupakan hal yang krusial topik penelitian. Namun, banyak pekerjaan di bidang data mining dibangun di atas adanya data yang berkualitas [7]. Yaitu, input ke algoritma penambangan data diasumsikan terdistribusi dengan baik, tidak mengandung nilai yang hilang atau salah di mana semua fitur

penting. Hal ini menyebabkan:

1. Menyamarkan pola yang berguna yang tersembunyi dalam data,
2. Kinerja rendah, dan
3. Kualitas buruk keluaran. Untuk memulai dengan upaya terfokus dalam persiapan data, edisi khusus ini termasuk dua belas makalah yang dipilih dari Lokakarya Internasional Pertama tentang Pembersihan dan Prapemrosesan Data.

3.2. Pentingnya Persiapan Data Pada Data Mining

Selama bertahun-tahun, telah ada kemajuan signifikan dalam penambangan data teknik. Kemajuan ini belum diimbangi dengan kemajuan serupa di persiapan data. Oleh karena itu, sekarang ada kebutuhan yang kuat untuk teknik baru dan alat otomatis yang akan dirancang yang secara signifikan dapat membantu kami dalam menyiapkan data yang berkualitas. Persiapan data bisa lebih memakan waktu daripada data penambangan, dan dapat menghadirkan tantangan yang setara, jika tidak lebih, daripada penambangan data [8]. Pada bagian ini, kami memperdebatkan pentingnya persiapan data pada tiga aspek:

1. Data dunia nyata tidak murni;
2. Penambangan berkinerja tinggi sistem membutuhkan data yang berkualitas; dan
3. Data berkualitas menghasilkan pola berkualitas tinggi.

Data real mungkin tidak lengkap, berisik, dan tidak konsisten, yang dapat menyamarkan pola yang berguna. Ini berhubungan dengan: Data tidak lengkap: tidak memiliki nilai atribut, tidak memiliki atribut tertentu minat, atau hanya berisi data agregat. Data bising: mengandung kesalahan atau outlier. Data tidak konsisten: mengandung perbedaan kode atau nama. Persiapan data menghasilkan kumpulan data yang lebih kecil dari yang asli, yang dapat secara signifikan meningkatkan efisiensi data mining. Tugas ini meliputi:

1. Memilih data yang relevan: pemilihan atribut (pemfilteran dan pembungkus metode), menghapus anomali, atau menghilangkan catatan duplikat.

2. Pengurangan data: pengambilan sampel atau pemilihan contoh.
3. Penyiapan data menghasilkan data yang berkualitas, yang mengarah pada pola-pola yang berkualitas.
4. Misalnya, kita dapat: Pulihkan data yang tidak lengkap: mengisi nilai yang terlewat, atau mengurangi kemenduaan. Memurnikan data: mengoreksi kesalahan, atau menghapus outlier (tidak biasa atau nilai luar biasa). Menyelesaikan konflik data: menggunakan pengetahuan domain atau keputusan ahli untuk menyelesaikan perbedaan

3.3. Algoritma C4.5

Data Mining adalah proses eksplorasi terhadap data yang besar untuk menemukan pola dalam pengambilan keputusan. Salah satu teknik dalam pengambilan keputusan adalah klasifikasi. Klasifikasi merupakan suatu teknik dalam data mining dengan menerapkan metode pohon keputusan untuk membentuk data, algoritma C4.5 merupakan algoritma yang dapat digunakan untuk mengklasifikasikan data dalam bentuk pohon [9]. Sistem yang telah dibangun menunjukkan hasil kinerja yang baik dan minim error dilihat dari sistem yang mampu membedakan trafik anomali dengan trafik normal. Aplikasi sistem persediaan data mining dapat mempermudah pengendalian persediaan di perusahaan untuk menekan biaya produksi.

Masalah persediaan merupakan salah satu masalah penting yang harus diselesaikan oleh perusahaan. Salah satu upaya untuk mengantisipasi masalah persediaan ini adalah dengan membangun sistem pengendalian persediaan untuk mengendalikan kelebihan atau kekurangan persediaan [10]. Jika perusahaan memiliki kelebihan persediaan maka banyak resiko yang harus diatasi seperti kemungkinan kerusakan barang, biaya perawatan, serta modal yang besar. Sebaliknya jika perusahaan kekurangan persediaan maka akan menimbulkan kekecewaan bagi pelanggan, hilangnya kesempatan untuk memperoleh keuntungan dan menimbulkan rasa kurang percaya dari pelanggan serta

beralihnya pelanggan ke produsen lain yang akan merugikan perusahaan itu sendiri. Permasalahan yang dihadapi perusahaan seperti penjelasan sebelumnya dapat diatasi dengan menerapkan metode data mining]. Data mining dimaksudkan untuk memberikan solusi nyata bagi para pengambil keputusan, untuk mengembangkan bisnisnya. Dan juga data mining dapat melakukan peramalan data Metode data mining yang dapat diterapkan seperti metode klasifikasi C4.5 [11]. Beberapa penelitian yang dilakukan melalui database transaksi penjualan dapat diperoleh berbagai informasi tentang kebiasaan atau perilaku konsumen. Misalnya dapat diketahui produk apa saja yang sering dibeli dan jarang dibeli konsumen dalam setiap transaksi, dan diketahui saat terjadi peningkatan pembelian oleh konsumen dan juga melakukan pembungkangan di pasar dengan berbagai promosi, jika perilaku konsumen dalam pengambilan keputusan pembuatannya dapat diketahui, maka perusahaan dapat merancang berbagai strategi pengendalian persediaan [12].

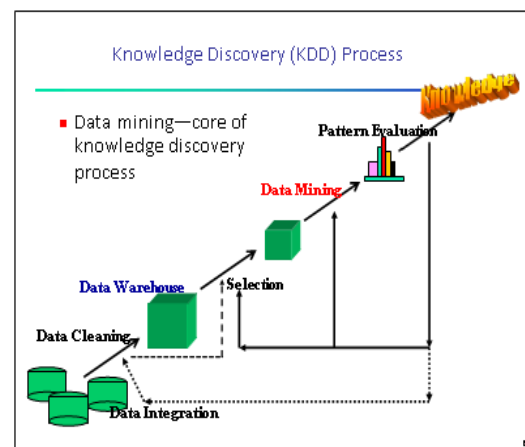
Data Mining adalah bidang ilmu multidisiplin, menggambarkan area kerja yang mencakup teknologi database, pembelajaran mesin, statistik, pengenalan pola, pencarian informasi, jaringan saraf tiruan, sistem berbasis pengetahuan, kecerdasan buatan, komputasi kinerja tinggi, dan visualisasi data. Data Mining didefinisikan sebagai data mining atau upaya untuk mengekstrak informasi berharga dan berguna pada database yang sangat besar. Hal terpenting dalam teknik data mining adalah aturan untuk menemukan pola frekuensi tinggi di antara set item set yang disebut fungsi Association Rules. Algoritma Klasifikasi C4.5 atau juga dikenal dengan algoritma pohon keputusan adalah metode klasifikasi dan prediksi yang sangat kuat dan terkenal [13]. Algoritma ini termasuk metode data mining, yaitu proses menemukan pola dengan menyaring sejumlah besar data menggunakan teknologi pengenalan pola.

Analisis data dan pendukung keputusan Data Mining [14]:

1. Analisis dan manajemen pasar

Pemasaran target,

2. Manajemen hubungan pelanggan (CRM), analisis keranjang pasar, penjualan silang, segmentasi pasar
3. Analisis dan manajemen risiko
4. Peramalan, retensi pelanggan, penjaminan emisi yang lebih baik, kontrol kualitas, analisis kompetitif
5. Deteksi penipuan dan deteksi pola yang tidak biasa (pencilan)
6. Aplikasi Lainnya
7. Penambangan teks (grup berita, email, dokumen) dan penambangan Web



Gambar 1. Knowledge Discovery (KDD) Process [15]

Jumlah data yang besar, Algoritma harus sangat skalabel untuk ditangani seperti data tera-byte, Data berdimensi tinggi, Micro-array mungkin memiliki puluhan ribu dimensi, Kompleksitas data yang tinggi, Aliran data dan data sensor, Data deret waktu, data temporal, data urutan, Data struktur, grafik, jejaring sosial, dan data multi-tautan.

Basis data heterogen dan basis data lama, Data spasial, spatiotemporal, multimedia, teks dan Web, Program perangkat lunak, simulasi ilmiah, Aplikasi baru dan canggih.

3.4. Tampilan Multi-Dimensi Data Mining

1. Data yang akan ditambang.
Relasional, gudang data, transaksional, aliran, berorientasi objek/relasional, aktif, spasial, deret waktu, teks, multimedia, heterogen, warisan, WWW
2. Pengetahuan untuk ditambang.

- Karakterisasi, diskriminasi, asosiasi, klasifikasi, pengelompokan, tren/deviasi, analisis outlier, dll. Beberapa / fungsi terintegrasi dan penambangan di berbagai level b [16]
3. Teknik yang digunakan.
Berorientasi database, gudang data (OLAP), pembelajaran mesin, statistik, visualisasi, dll.
 4. Aplikasi disesuaikan.
Ritel, telekomunikasi, perbankan, analisis penipuan, penambangan bio-data, analisis pasar saham, penambangan teks, penambangan web, dll.

3.5. Fungsi Data Mining

Fungsi deskripsi dalam data mining adalah sebuah fungsi untuk memahami lebih jauh tentang data yang diamati. Dengan melakukan sebuah proses diharap bisa mengetahui perilaku dari sebuah data tersebut. Data tersebut itulah yang nantinya dapat digunakan untuk mengetahui karakteristik dari data yang dimaksud. Dengan menggunakan fungsi descriptive data mining, maka nantinya bisa menemukan pola tertentu yang tersembunyi dalam sebuah data. Dengan kata lain jika pola yang berulang dan bernilai itulah karakteristik sebuah data bisa diketahui.

Sedangkan fungsi prediksi merupakan sebuah fungsi bagaimana sebuah proses nantinya akan menemukan pola tertentu dari suatu data. Pola-pola tersebut dapat diketahui dari berbagai variabel yang ada pada data. Ketika sudah menemukan pola, maka pola yang didapat tersebut bisa digunakan untuk memprediksi variabel lain yang belum diketahui nilai ataupun jenisnya. Oleh karena itu, fungsi ini memudahkan dan menguntungkan bagi siapapun yang memerlukan prediksi yang akurat.

4. KESIMPULAN

Berkat kemajuan komputer dan teknologi pengambilan data, kumpulan data yang sangat besar-berisi gigabyte atau bahkan terabyte data olah dan sedang dikumpulkan.

Kumpulan data ini mengandung informasi yang berpotensi berharga. Triknya

adalah mengekstrak informasi berharga itu dari kumpulan angka-angka yang tidak menarik di sekitarnya, sehingga pemilik data dapat memanfaatkannya. Penambangan data adalah disiplin baru yang berusaha melakukan hal itu: dengan memilah-milah basis data ini, meringkasnya, dan menemukan pola.

Penambangan data tidak boleh dilihat sebagai latihan satu kali yang sederhana. Kumpulan data yang sangat besar dapat dianalisis dan diperiksa dengan cara yang tidak terbatas. Seiring berjalannya waktu, jenis struktur dan pola baru dapat menarik minat, dan mungkin perlu dicari dalam data. Data mining, untuk alasan yang baik, baru-baru ini menarik banyak perhatian: ini adalah teknologi baru, mengatasi masalah baru, dengan potensi besar untuk penemuan komersial dan ilmiah yang berharga. Namun, kita tidak boleh mengharapkannya untuk memberikan jawaban atas semua pertanyaan. Seperti semua proses penemuan, penambangan data yang sukses memiliki unsur kebetulan. Meskipun penambangan data menyediakan alat yang berguna, itu tidak berarti bahwa itu pasti akan mengarah pada hasil yang penting, menarik, atau berharga. Kita harus berhati-hati untuk melebih-lebihkan kemungkinan hasil.

5. SARAN

Semua hasil yang telah dicapai saat ini, tentunya masih banyak kekurangan yang harus diperbaiki atau ditambah. Saran-saran terhadap penggunaan algoritma C4.5 adalah sebagai berikut :

1. Perlu adanya data yang lebih kompleks agar keakuratan prediksi semakin tinggi.
2. Perlu adanya pengembangan lagi agar pohon keputusan ini bisa memprediksi lebih tepat, terukur dan akurat.

DAFTAR PUSTAKA

- [1] A. a. n. Mostafa, "Review of data mining concept and its techniques," *Innov. Technol.*, vol. 9, no. 4, 2018.
- [2] A. Sumiah, D. Irawan, F. Ilmu, K. Universitas, and J. N. Informatika, "Perancangan Data Warehouse Dan

- Penerapan Data Mining Untuk Mendukung Sistem Informasi Akademik,” *J. Information Syst.*, vol. 8, 2014.
- [3] G. Singer, M. Golan, N. Rabin, and D. Kleper, “Evaluation of the effect of learning disabilities and accommodations on the prediction of the stability of academic behaviour of undergraduate engineering students using decision trees,” *Eur. J. Eng. Educ.*, vol. 45, no. 4, pp. 614–630, Jul. 2020, doi: 10.1080/03043797.2019.1677560.
- [4] P. Aspers and U. Corte, “What is Qualitative in Qualitative Research,” *Qual. Sociol.*, vol. 42, no. 2, pp. 139–160, Jun. 2019, doi: 10.1007/s11133-019-9413-7.
- [5] N. Mirantika, “Penerapan Algoritma K-Means Clustering Untuk Pengelompokan Penyebaran Covid-19 di Provinsi Jawa Barat,” *NUANSA Inform.*, vol. 15, no. 2, 2021, doi: 10.25134/nuansa.v15i2.4321.
- [6] A. P. Joshi and B. V. Patel, “Data Preprocessing: The Techniques for Preparing Clean and Quality Data for Data Analytics Process,” *Orient. J. Comput. Sci. Technol.*, vol. 13, no. 0203, pp. 78–81, Jan. 2021, doi: 10.13005/ojctst13.0203.03.
- [7] J. Yang *et al.*, “Brief introduction of medical database and data mining technology in big data era,” *J. Evid. Based. Med.*, vol. 13, no. 1, pp. 57–69, Feb. 2020, doi: 10.1111/jebm.12373.
- [8] J. Goncalves *et al.*, “Li Yan *et al.* reply,” *Nat. Mach. Intell.*, vol. 3, no. 1, pp. 28–32, Jan. 2021, doi: 10.1038/s42256-020-00251-5.
- [9] R. Rahim *et al.*, “C4.5 Classification Data Mining for Inventory Control,” *Int. J. Eng. Technol.*, vol. 7, no. 2.3, p. 68, Mar. 2018, doi: 10.14419/ijet.v7i2.3.12618.
- [10] O. Abdolazimi, D. Shishebori, F. Goodarzian, P. Ghasemi, and A. Appolloni, “Designing a new mathematical model based on ABC analysis for inventory control problem: A real case study,” *RAIRO - Oper. Res.*, vol. 55, no. 4, pp. 2309–2335, Jul. 2021, doi: 10.1051/ro/2021104.
- [11] P. Wang, T. Weise, and R. Chiong, “Novel evolutionary algorithms for supervised classification problems: an experimental study,” *Evol. Intell.*, vol. 4, no. 1, pp. 3–16, Mar. 2011, doi: 10.1007/s12065-010-0047-7.
- [12] G. Bruno, E. Traini, and F. Lombardi, “A Knowledge-Based System for Collecting and Integrating Production Information,” in *IFIP Advances in Information and Communication Technology*, 2019, vol. 568, doi: 10.1007/978-3-030-28464-0_15.
- [13] M. V. Sebt, E. Komijani, and S. S. Ghasemi, “Implementing a data mining solution approach to identify the valuable customers for facilitating electronic banking,” *Int. J. Interact. Mob. Technol.*, vol. 14, no. 15, 2020, doi: 10.3991/IJIM.V14I15.16127.
- [14] E. Sugiyarti, K. A. Jasmi, B. Basiron, M. Huda, K. Shankar, and A. Maselena, “Decision support system of scholarship grantee selection using data mining,” *Int. J. Pure Appl. Math.*, vol. 119, no. 15, 2018, doi: 10.5772/47788.
- [15] A. BaniMustafa and N. Hardy, “A Scientific Knowledge Discovery and Data Mining Process Model for Metabolomics,” *IEEE Access*, 2020, doi: 10.1109/ACCESS.2020.3039064.
- [16] B. Budiman, “Perbandingan Algoritma Klasifikasi Data Mining untuk Penelusuran Minat Calon Mahasiswa Baru,” *NUANSA Inform.*, vol. 15, no. 2, 2021, doi: 10.25134/nuansa.v15i2.4162.