

PELATIHAN DAN SIMULASI DATA MINING PADA SMA NEGERI 3 SEMARANG

Farrikh AL Zami¹, Sendi Novianto², Asih Rohmani³, Sri Winarno⁴,
Indra Gamayanto⁵, Titien Suhartini Sukamto⁶, Sasono Wibowo⁷,
Ramadhan Rakhmat Sani⁸, Setyo Budi^{9*}

^{1,3,4,5,6,7,8,9}Program Studi Sistem Informasi, Universitas Dian Nuswantoro, Semarang,
Indonesia

²Program Studi Teknik Informatika, Universitas Dian Nuswantoro, Semarang,
Indonesia

*setyobudi@dsn.dinus.ac.id, Jl. Nakula 1 No.5-11, 50131, Semarang, Indonesia

Kata Kunci: *Data Mining, Studi Kasus, Penerapan, Simulasi, Data.*

Abstrak - SMA Negeri 3 Semarang adalah salah satu dari beberapa SMA terbaik di kota Semarang yang Siswa/i serta Gurunya mempunyai keahlian tinggi. Unsur utama yang dapat menaikkan keahlian SMA tersebut adalah telah dilakukannya kalaborasi dengan perguruan tinggi dalam kegiatan pelatihan dan peningkatan ilmu pengetahuan dari berbagai bidang, agar supaya dapat bersaing di era globalisasi. Sekarang ini menguatkan keahlian Siswa/i dan Guru merupakan hal utama, karena nantinya akan dapat bersaing dengan negara maju lainnya. Untuk menguatkan keahlian tersebut maka dilakukan pelatihan dan simulasi data *mining* di SMA Negeri 3 Semarang. Diawali dengan survei, kemudian dianalisis dan akhirnya dilaksanakan kegiatan pelatihan data *mining*. Hasil dari pengabdian masyarakat data *mining* ini adalah Siswa dan Guru mampu memahami data *mining*, contoh dan penerapannya, sehingga Siswa dan Guru memiliki pengetahuan khusus data *mining* yang berguna untuk meningkatkan kompetensi dan persaingan global.

Pendahuluan

SMA Negeri 3 Semarang adalah salah satu dari beberapa SMA terbaik di kota Semarang yang Siswa/i serta Gurunya mempunyai keahlian tinggi. Unsur utama yang dapat menaikkan keahlian SMA tersebut adalah dilakukannya kalaborasi dengan perguruan tinggi dalam kegiatan pelatihan dan peningkatan ilmu pengetahuan dari berbagai bidang. Sekarang ini menguatkan keahlian

Siswa/i dan Guru merupakan hal utama, karena untuk membekali mereka didalam bersaing dengan negara lain. Untuk menguatkan keahlian tersebut, perlu dilakukan pelatihan dan simulasi data *mining* di SMA Negeri 3 Semarang. Metode yang dilakukan adalah melakukan survei terlebih dahulu sehingga ditemukan permasalahan, kemudian dianalisis dan akhirnya diputuskan untuk diadakan kegiatan pelatihan dan simulasi data *mining*. Hasil dari pengabdian masyarakat ini adalah Sumber Daya Manusia (SDM) dalam hal ini Siswa/i beserta Guru akan mampu memahami data *mining*, contoh dan penerapannya, sehingga Siswa dan Guru memiliki pengetahuan khusus tentang data *mining*. Untuk mendukung pengembangan SDM yang berbasis teknologi informasi (*Human technology*), diperlukan antara lain: (1) Pemahaman tentang dasar-dasar data *mining* dan dampaknya bagi kehidupan manusia, pada bagian ini Siswa/i dan Guru mampu memahami apa yang harus diketahui tentang data *mining* secara mendalam, (2) Data *mining*, pada bagian ini Siswa/I dan Guru dapat memahami bagaimana data *mining* dapat memberikan keuntungan besar bagi diri sendiri dan orang lain, (3) Pemahaman yang mendalam bagaimana data *mining* dapat diterapkan disertai contoh studi kasus, (4) Bagaimana implementasi data *mining* dan bagaimana cara mengatasi kendala-kendala yang ada, (5) Bagaimana data *mining* dapat memiliki pengaruh yang signifikan terhadap perkembangan dan peningkatan taraf hidup orang banyak.

Dari hasil analisis tim pengabdian masyarakat, ditemukan permasalahan yang di hadapai mitra saat ini, yaitu membutuhkan pengetahuan untuk meningkatkan keahlian Siswa/i beserta Guru di era berbasis data seperti saat ini. Dimana permasalahan tersebut mencakup dibutuhkannya pengetahuan yang lebih luas sebelum Siswa/i lulus yaitu harus dibekali dengan pengetahuan di bidang teknologi informasi, bidang sosial media, industri kreatif game, penggalian data, bisnis *online*, mengawali bisnis baru dan sebagainya. Tentunya salah satu cara untuk membekalinya adalah dengan diadakannya pelatihan dan simulasi tentang ilmu dan pengetahuan yang berkaitan dengan bidang yang sudah disebutkan diatas. Selanjutnya target dari pengabdian masyarakat ini adalah: (1) Meningkatkan keahlian Siswa/i dan Guru di bidang ilmu data *mining*, (2) Pemahaman yang matang tentang data *science*, (3) Mampu mengimplementasikan data *mining* ke dalam dunia nyata. Kenapa pelatihan ini ditekankan pada data *mining*, karena data *mining* merupakan ilmu yang dapat digunakan untuk menganalisis data yang sangat besar dan dapat mengidentifikasi pengetahuan yang berharga untuk manusia [1].

Metode Pelaksanaan

Pelaksanaan pelatihan dan simulasi data *mining* di SMA Negeri 3 Semarang dilaksanakan dengan beberapa tahapan yaitu: 1) Diskusi, wawancara dan survei di SMA Negeri 3 Semarang; 2) Analisis permasalahan; 3) Pelatihan data *mining* dan; 4) Kompetensi data *mining*. Hal ini bisa dilihat di gambar 1.



Gambar 1. Tahapan pelatihan dan pengabdian masyarakat

Pada gambar 1, menjelaskan bahwa pertama kali yang dilakukan oleh tim pengabdian adalah melakukan diskusi, wawancara dan analisis permasalahan sehingga pengabdian masyarakat yang akan di lakukan dapat tepat pada sasaran. Permasalahan yang ditemukan adalah dibutuhkannya pengetahuan untuk meningkatkan keahlian Siswa/i beserta Guru di era berbasis data seperti saat ini. Untuk peningkatan keahlian tersebut maka disepakati dilaksanakan kegiatan pelatihan dan simulasi data *mining* sampai dengan pendampingan secara berkala sehingga sasaran pelatihan dapat tercapai. Detail metode yang digunakan adalah: (1) Penjelasan data *mining* secara umum dan khusus, (2) Penjelasan data *mining* dan penerapan serta implementasi di berbagai sektor. Pelatihan ini dilakukan terangkum pada dua aktivitas yang terencana yaitu: Aktivitas 1: Mempersiapkan pendampingan, melaksanakan kegiatan pendampingan dalam menyampaikan konsep data *mining* dan dampaknya; Aktivitas 2: Pendampingan teknologi informasi, di tahap ini, Tim Pengabdian Masyarakat bersama SMA Negeri 3 Semarang mempelajari dan cara mengimplementasikan data *mining*.

Hasil dan Pembahasan

Kegiatan Pelatihan dan Pengabdian Masyarakat

Langkah awal yang dilakukan tim adalah diskusi, wawancara dan melihat lokasi di SMA Negeri 3 Semarang. Gambar 3 menjelaskan, diskusi yang dilakukan tim bersama pihak SMA Negeri 3 Semarang untuk mengetahui masalah tengan peningkatan keahlian. Hasil dari diskusi dan analisis tim, di sepakati diadakannya pelatihan dan simulasi data *mining*.



Gambar 2. Diskusi dengan Guru SMA Negeri 3 Semarang

Hasil Pelatihan dan Pengabdian Masyarakat & Implementasi



Gambar 3. Contoh Materi Data Mining

Data Mining

Tujuan dari data mining adalah menemukan pola dalam data sehingga pengetahuan ini dapat diterapkan pada pemecahan masalah. Analitik data mining yang terintegrasi dengan visualisasi yang kuat menghadirkan jalur baru menuju penemuan pengetahuan. Sistem data mining dapat secara otomatis menemukan dan menunjukkan pola baru yang akan menghasilkan wawasan baru [2]. Contoh dari ini mungkin menentukan korelasi di antara atribut, membedakan antara sub kumpulan data dengan karakteristik yang berbeda, dan menyimpulkan probabilitas peristiwa masa depan dari data historis [3][4].

Dalam kueri database biasa atau pemrosesan analitik *online* (OLAP), pengguna harus menentukan secara langsung hubungan apa pun antara elemen data. Jumlah data yang diambil mungkin sangat besar sehingga menghalangi analisis dengan cara selain penggalian data. Data semacam itu, setelah diubah dengan benar, sering disimpan di gudang data [5].

Data Mining Methods

Data *mining* menggabungkan pengujian hipotesis dan penemuan berdasarkan data. Dalam pengujian hipotesis, penyidik menguji ide terhadap sekumpulan data untuk mengkonfirmasi atau menolak validitasnya. Dalam beberapa kasus, data itu sendiri dapat mendorong penemuan. Dalam

penemuan, penyidik menarik kesimpulan dari data, memungkinkan data itu sendiri untuk memberi kesimpulan. Seringkali masalah data mining diselesaikan dengan menggunakan perpaduan kedua metode. Misalnya, kesimpulan dapat menimbulkan hipotesis baru yang dapat diuji, dan dikonfirmasi atau ditolak. Penambangan data adalah tempat statistik dan pembelajaran mesin bertemu [6].

Analytical Data Mining Algorithms

Algoritma penambangan data analitik secara otomatis membangun model dari data. Dua kelompok algoritme pemodelan umumnya digunakan — diawasi dan tidak diawasi. Tugas pemodelan prediktif, di mana tujuannya adalah untuk memprediksi nilai satu kolom berdasarkan nilai kolom lainnya, disebut tugas yang diawasi [7]. Tugas-tugas ini mirip dengan pengawasan seorang Guru yang memberi Anda jawaban yang benar untuk pertanyaan, untuk mengajari kita.

Supervised Modeling

Dalam pemodelan terawasi, ada atribut khusus yang disebut “label” yang ingin kita prediksi [8]. Dengan mengkode hubungan antara label dan atribut lainnya, model dapat membuat prediksi tentang data baru yang tidak berlabel [9]. Selain itu, dengan memvisualisasikan model itu sendiri, kita dapat memperoleh wawasan tentang hubungan antara label dan atribut lainnya [10]. Misalnya, jika pelanggan telah meninggalkan perusahaan kita (biasanya disebut atrisi atau churn), kita dapat membuat model yang tidak hanya akan memprediksi pelanggan mana yang cenderung keluar, tetapi juga membantu kita memahami alasan dan pola yang mengarah pada perilaku ini.

Klasifikasi

Klasifikasi adalah tugas untuk menetapkan nilai label diskrit ke *record* yang tidak berlabel. Pengklasifikasi adalah model yang memprediksi satu atribut dari sekumpulan data ketika diberikan atribut lain [11]. Dengan demikian, rekaman dibagi menjadi grup yang telah ditentukan sebelumnya. Misalnya, klasifikasi sederhana mungkin mengelompokkan catatan tagihan pelanggan ke dalam dua kelas spesifik: mereka yang membayar tagihan dalam 60 hari, dan mereka yang membutuhkan waktu lebih dari 60 hari untuk membayar. Contoh klasifikasi data lebih lanjut mungkin membagi pelanggan berdasarkan jenis kelamin atau pendapatan. Pengklasifikasi juga dapat memprediksi probabilitas bahwa label akan mengambil nilai tertentu. Misalnya, probabilitas orang tersebut akan membayar tagihannya dalam 60 hari dapat dihitung.

Regresi

Regresi adalah tugas pemodelan yang diawasi yang mirip dengan klasifikasi, kecuali labelnya tidak terpisah. Misalnya, memprediksi gaji atau

harga saham adalah regresi, sedangkan memprediksi apakah gaji dalam kisaran tertentu atau apakah saham akan naik atau turun adalah tugas klasifikasi.

Menilai Akurasi Model

Model prediktif jarang sekali sempurna, oleh karena itu memperkirakan keakuratannya adalah bagian penting dari proses data mining. Alat yang digunakan untuk mengukur akurasi bergantung pada jenis model. Pengklasifikasi biasanya dievaluasi menurut tingkat kesalahannya. Ukuran yang paling umum adalah kesalahan klasifikasi, atau proporsi catatan yang salah diklasifikasikan. Saat menilai akurasi model, penting untuk mengujinya pada data yang tidak digunakan dalam membangun model. MineSet menyediakan sejumlah metode untuk mengevaluasi kesalahan.

Pemodelan Tanpa Pengawasan

Dalam pemodelan tanpa pengawasan, tujuannya adalah untuk menemukan aturan dan segmen data yang berperilaku serupa (*cluster*). Pemodelan tanpa pengawasan adalah tugas deskriptif, bukan tugas prediksi. Model tidak dapat digunakan secara langsung untuk membuat prediksi, oleh karena itu tidak perlu menyisihkan sebagian data sebagai set pelatihan untuk membangun pengklasifikasi. Dua tugas pemodelan tanpa pengawasan yang paling umum adalah asosiasi dan pengelompokan.

Asosiasi

Untuk menghasilkan asosiasi, tugasnya adalah menentukan aturan implikasi antara atribut data sehingga A menyiratkan B. Asosiasi digunakan untuk menemukan pengelompokan afinitas yang menemukan item apa yang biasanya dibeli dengan orang lain. Pengelompokan afinitas klasik adalah analisis keranjang pasar, yang memprediksi frekuensi pembelian barang tertentu pada waktu yang sama. Misalnya, mengetahui bahwa makanan bayi menyiratkan probabilitas yang lebih tinggi bahwa pelanggan akan membeli rokok rendah tar daripada rokok biasa dapat membantu toko mengatur rak mereka secara berbeda. Asosiasi dapat dilihat di MineSet menggunakan *Rules Visualizer*.

Clustering

Algoritme pengelompokan membagi data menjadi beberapa kelompok catatan, atau kelompok, yang memiliki karakteristik serupa. Misalnya, perusahaan asuransi kesehatan mungkin menemukan bahwa karakteristik ini menentukan segmen: usia 20 hingga 45 tahun, pekerja teknis, kurang dari dua anak, penggemar fiksi ilmiah televisi, dan pendapatan yang dapat dibuang sebesar \$ 5.000 hingga \$ 10.000 per tahun. Segmen tersebut kemudian dapat

ditargetkan secara lebih efektif dengan paket asuransi kesehatan yang sesuai untuk orang-orang ini, dengan menggunakan iklan televisi dalam episode fiksi ilmiah baru.

Visualisasi Data

Algoritma penambangan data analitik dapat dilengkapi dengan teknik visualisasi data yang memanfaatkan kemampuan pengenalan pola otak manusia yang menakjubkan. *Visualizer* MineSet berikut tersedia: 1) *Visualizer* Peta; 2) *Scatter Visualizer*; 3) *Splat Visualizer*; 4) *Tree Visualizer*.

Manfaat Data Mining

Ada beberapa bidang yang membuktikan bahwa data *mining* dapat mengatasi masalah, antara lain:

- 1) Pengelolaan hubungan pelanggan
 - a) Dengan tujuannya untuk membangun hubungan satu lawan satu dengan pelanggan dengan mengembangkan pemahaman yang mendalam tentang kebutuhan dan keinginan mereka, data mining bisa sangat berguna. Dengan semua data yang dihasilkan dari berbagai acara (pertanyaan produk, penjualan, ulasan produk), ada banyak cara berbeda untuk data mining dapat memberikan lebih banyak wawasan [12].
 - b) Identifikasi pembeli / responden yang paling mungkin dari produk dan layanan baru [13].
 - c) Pahami akar penyebab atrisi pelanggan untuk meningkatkan retensi pelanggan.
 - d) Temukan asosiasi varian waktu antara produk dan layanan untuk memaksimalkan penjualan dan nilai pelanggan.
 - e) Identifikasi pelanggan yang paling menguntungkan, dan kebutuhan preferensial mereka untuk memperkuat hubungan dan memaksimalkan penjualan.
- 2) Industri ritel
 - a) Memprediksi volume penjualan yang akurat pada tingkat inventaris tertentu.
 - b) Identifikasi hubungan penjualan antara berbagai jenis produk (analisis keranjang pasar).
 - c) Perkiraan tingkat konsumsi berbagai jenis produk (berdasarkan kondisi musiman dan lingkungan) untuk mengoptimalkan logistik dan karenanya memaksimalkan pendapatan.
 - d) Temukan pola yang menarik dalam pergerakan produk, terutama produk dengan umur simpan yang pendek, dalam rantai pasokan dengan menganalisis data sensorik dan RFID.
- 3) Manufaktur dan produksi
 - a) Prediksi kegagalan mesin sebelum terjadi dengan menggunakan data sensorik, yang akan memungkinkan pemeliharaan berbasis kondisi [14].

- b) Identifikasi kesamaan dan anomali dalam sistem produksi untuk mengoptimalkan kapasitas produksi.
 - c) Temukan pola baru untuk mengidentifikasi dan meningkatkan kualitas produk [15]
- 4) Industri perjalanan (maskapai atau hotel)
- a) Berguna untuk memprediksi penjualan berbagai layanan (jenis kursi di pesawat, jenis kamar hotel) untuk menentukan harga layanan secara optimal guna memaksimalkan pendapatan sebagai fungsi manajemen hasil.
 - b) Perkiraan permintaan di lokasi yang berbeda untuk mengalokasikan sumber daya organisasi yang terbatas dengan lebih baik.
 - c) Identifikasi pelanggan yang paling menguntungkan dan berikan mereka layanan yang dipersonalisasi untuk mempertahankan bisnis berulang mereka.
 - d) Pertahankan karyawan yang berharga dengan mengidentifikasi dan bertindak berdasarkan akar penyebab atrisi.

Kesimpulan

Setelah kami melakukan pelatihan dan pengabdian masyarakat di SMA Negeri 3 Semarang, kesimpulan yang kami dapat adalah perlu adanya peningkatan kemampuan dan pemahaman Siswa/i dan Guru dalam bidang data mining dan bagaimana menjadi seorang data *science*, mampu mengimplementasikan data mining dalam kehidupan sehari-hari dan mampu mengimplementasikan data mining pada bidang startup. Untuk mencapai itu semua maka saran yang dapat kami berikan kepada SMA Negeri 3 Semarang adalah secara *continue* kegiatan pengabdian masyarakat tetap berjalan dalam jangka waktu yang panjang sehingga kompetensi data mining di Siswa/i serta Guru di SMA Negeri 3 Semarang selalu meningkat

Ucapan Terima Kasih

Tidak ada kata lain yang dapat kami sampaikan kecuali ucapan terima kasih sebesar-besarnya kepada Siswa/i dan Guru SMA Negeri 3 Semarang yang telah memberikan kesempatan dan bekerjasama demi terlaksananya kegiatan pelatihan dan simulasi data *mining* yang sudah terlaksana dengan baik. Harapan kami pelatihan ini dapat meningkatkan keahlian Siswa/i dalam menghadapi era globalisasi.

Referensi

- [1] K. Vadim, "Overview of different approaches to solving problems of data mining," *Procedia Comput. Sci.*, vol. 123, pp. 234–239, 2018, doi: 10.1016/j.procs.2018.01.036.
- [2] V. Plotnikova, M. Dumas, and F. Milani, "Adaptations of data mining methodologies: A systematic literature review," *PeerJ Comput. Sci.*, vol. 6, pp. 1–43, 2020, doi: 10.7717/PEERJ-CS.267.

- [3] G. Zhu, "Precision Retail Marketing Strategy Based on Digital Marketing Model," *Sci. J. Bus. Manag.*, vol. 7, no. 1, p. 33, 2019, doi: 10.11648/j.sjbm.20190701.15.
- [4] M. K. J. Pei, Jiawei Han, *Data mining: Data mining concepts and techniques*. 2012.
- [5] A. Novikov, "PyClustering: Data Mining Library," *J. Open Source Softw.*, vol. 4, no. 36, p. 1230, 2019, doi: 10.21105/joss.01230.
- [6] S. H. Bhojani and N. Bhatt, "Data Mining Techniques and Trends – A Review Commerce Computer Science Data Mining Techniques and Trends – A Review Assistant Professor-Anand Agriculture University-Anand - Gujarat Assistant Professor – RK University , Rajkot - Gujarat," no. May 2016, pp. 1–4, 2018.
- [7] T. Marnoto, "Drying of Rosella (Hibiscus sabdariffa) Flower Petals using Solar Dryer with Double Glass Cover Collector," *Int. J. Sci. Eng.*, vol. 7, no. 2, pp. 155–160, 2014, doi: 10.12777/ijse.7.2.150-154.
- [8] S. A. N. Alexandropoulos, S. B. Kotsiantis, and M. N. Vrahatis, *Data preprocessing in predictive data mining*, vol. 34, no. January. 2019.
- [9] M. Sharma, S. Sharma, and G. Singh, "Performance analysis of statistical and supervised learning techniques in stock data mining," *Data*, vol. 3, no. 4, pp. 1–16, 2018, doi: 10.3390/data3040054.
- [10] S. Rezig, Z. Achour, and N. Rezg, "Using data mining methods for predicting sequential maintenance activities," *Appl. Sci.*, vol. 8, no. 11, 2018, doi: 10.3390/app8112184.
- [11] L. Colonna, "A Taxonomy and Classification of Data Mining," *Sci. Technol. Law Rev.*, vol. 16, no. 2, pp. 309–369, 2013.
- [12] M. R. Ayyagari, "A Framework for Analytical CRM Assessments Challenges and Recommendations," *Int. J. Bus. Soc. Sci.*, vol. 10, no. 6, pp. 5–13, 2019, doi: 10.30845/ijbss.v10n6p2.
- [13] L. Kharb, "Data Mining- A Distinctive approach to CRM," *Ijsrnsc*, vol. 7, no. 2, 2019.
- [14] G. Schuh *et al.*, "Data mining definitions and applications for the management of production complexity," *Procedia CIRP*, vol. 81, pp. 874–879, 2019, doi: 10.1016/j.procir.2019.03.217.
- [15] S. Chen, X. Li, R. Liu, and S. Zeng, "Extension data mining method for improving product manufacturing quality," *Procedia Comput. Sci.*, vol. 162, no. Itqm 2019, pp. 146–155, 2019, doi: 10.1016/j.procs.2019.11.270.

