

Diterima : February 01, 2021
Disetujui : February 05, 2021
Diterbitkan: February 24, 2021

**Conference on Management, Business,
Innovation, Education and Social Science**
<https://journal.uib.ac.id/index.php/combrates>

Perancangan Model Trend Prediksi Penerimaan Mahasiswa Baru Dengan Menggunakan Analisis Data (Studi Kasus: Universitas Xyz)

Tony Wibowo¹, Wiranto Mahendrako²

Email: tony.wibowo@uib.ac.id, 1731002.wiranto@uib.edu

^{1,2}Sistem Informasi, Universitas Internasional Batam, Batam, Indonesia

Abstrak

Analisis data penerimaan mahasiswa baru sangat penting dalam memformulasi strategi pemasaran yang tepat. Penulis berniat melakukan analisis data supaya dapat memperoleh gambaran dari mahasiswa dengan studi kasus pada Universitas XYZ. Dengan menggunakan analisis *data mining* dengan pendekatan *clustering* data dengan algoritma K-means terhadap data penerimaan mahasiswa baru dalam Universitas XYZ, kami menemukan 3 *cluster* terbesar yang mampu menjelaskan profil mahasiswa baru pada Universitas XYZ. Dari hasil tersebut Penulis memberikan rekomendasi terhadap strategi pemasaran yang akan datang.

Kata Kunci

Data mining, clustering, K-means

Abstract

Analyzing data of new students' admission is crucial in forming an accurate marketing strategy. We intend to perform data analysis to get an overview of students in the case study on XYZ University. By using data mining with the clustering approach using the K-means algorithm on the new students' admission data of XYZ University, we have discovered 3 largest clusters which can describe the profile of new students in XYZ University. From the results we have provided recommendations for marketing strategies in the future.

Keywords

Data mining, clustering, K-means

Pendahuluan

Pada masa kini dimana teknologi informasi sudah semakin berkembang dan berbagai perusahaan dan/atau organisasi saling bersaing. Data tidak hanya sekedar sebagai keterangan atau fakta dari suatu hal tertentu. Data dalam jumlah besar dapat diolah untuk membantu perusahaan atau organisasi mencapai sasaran mereka. Misalnya, data penjualan dari suatu perusahaan deterjen dilakukan analisa dengan metode *Online Analytical Processing* (OLAP) untuk memprediksi jumlah penjualan produk dalam negara yang berbeda-beda berdasarkan sales channel yang digunakan perusahaan, dalam penelitian oleh Akbar, Yuliani, Mawaddah, & Ardhana (2017). Proses untuk menggali informasi dan pola dari data dikenal dengan istilah *Data Mining*.

Menurut Mardi (2017), tujuan dari *data mining* yaitu untuk menemukan suatu kebiasaan atau pola tertentu dari sekumpulan data, yang mungkin bermanfaat untuk diketahui. Berdasarkan dari kesimpulan yang diperoleh dari *data mining*, suatu perusahaan atau organisasi akan mampu membuat keputusan yang baik merubah pola kerja meningkatkan kemajuan perusahaan atau organisasi, misalnya pada penerapan *data mining* di PT. Panen Lestari Indonesia (Syahril, Erwansyah, & Yetri, 2020). Pada penelitian tersebut, dengan *data mining* perusahaan dapat membuat keputusan yang baik dalam mengatur stok barang serta strategi penjualan dari perusahaan. Hal tersebut bisa dicapai dengan penggunaan pemodelan data yang dihasilkan oleh *data mining*.

Salah satu contoh *data mining* yaitu pada apotek Kimia Farma cabang Medan. Proses *data mining* dengan menggunakan algoritma Apriori untuk memperoleh prediksi pemesanan obat dari pelanggan. Dengan model prediksi tersebut, perusahaan dapat menyediakan obat-obat tertentu dalam jumlah yang sesuai dengan kebutuhan pelanggan dan tidak terjadi kelebihan ketersediaan obat (Arifin, 2020). *Data mining* juga dipergunakan dalam penelitian oleh Chasanah & Widiyono (2017) di STMIK Widya Pratama Pekalongan. *Data mining* dilakukan dengan menggunakan algoritma K-Means untuk menemukan informasi terkait kelompok-kelompok mahasiswa yang mendaftar pada tahun sebelumnya. Hal ini dapat digunakan untuk menentukan strategi promosi penerimaan mahasiswa baru pada STMIK Widya Pratama Pekalongan. Selain itu, *data mining* juga dilakukan dalam Rumah Sakit Umum Daerah Pekanbaru untuk membantu dalam menentukan strategi pengendalian dan perencanaan stok obat di rumah sakit tersebut (Gustientiedina, Adiya, & Desnelita, 2019).

Universitas XYZ adalah salah satu universitas yang dikenal di daerah A dan sekitarnya. Universitas XYZ memiliki keunggulan kompetitif dengan memiliki program kerjasama internasional dan secara konsisten menjadi universitas swasta terbaik di lingkungan LLDIKTI wilayah X. Universitas XYZ telah menerima banyak mahasiswa dari berbagai wilayah dan tidak hanya dari daerah A dan sekitarnya, namun ke daerah yang jauh secara geografis. Pada saat ini belum ada didefinisikan secara definitif profil calon mahasiswa di Universitas XYZ karena belum pernah dilakukan pengujian secara data pola penerimaan mahasiswa dan profil mahasiswa yang masuk. Dengan tidak adanya analisis data tersebut sangat sulit untuk menyusun strategi pemasaran yang tepat sasaran. Penulis tertarik untuk memperoleh gambaran besar dan prediksi mahasiswa yang mendaftar dan diterima di Universitas XYZ.

Tinjauan Pustaka

Berikut merupakan jurnal-jurnal penelitian yang Penulis gunakan sebagai tinjauan pustaka untuk penelitian ini.

Jurnal penelitian yang dikembangkan oleh Yunita (2018) berjudul "Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Pada Penerimaan Mahasiswa Baru". Penelitian tersebut bertujuan untuk melakukan pengelompokan data penerimaan mahasiswa baru pada Universitas Islam Indragiri dengan melakukan *data mining*. Peneliti menggunakan algoritma K-means dan teknik *clustering*, serta aplikasi yang digunakan yaitu RapidMiner 5.3. Hasilnya terbentuk tiga buah *cluster*/kelompok, dan telah dapat membantu Universitas Islam Indragiri dalam menentukan strategi promosi program-program studi di universitas tersebut.

Jurnal penelitian yang dikembangkan oleh Lestari (2019) berjudul "Clustering Data Mahasiswa Menggunakan Algoritma K-Means Untuk Menunjang Strategi Promosi (Studi Kasus: STMIK Bina Bangsa Kendari)". Metode yang digunakan oleh peneliti dalam penelitian tersebut yaitu dengan melakukan *data mining*, dengan menggunakan teknik *clustering* dan algoritma K-

means, yang menghasilkan pengelompokan mahasiswa berdasarkan data yang digunakan sehingga dapat membantu STMIK Bina Bangsa Kendari menentukan strategi promosi yang tepat sasaran.

Jurnal penelitian yang dikembangkan oleh Abriyanto & Damastuti (2019) berjudul "Segmentasi Mahasiswa Dengan Unsupervised Algoritma Guna Membangun Strategi Marketing Penerimaan Mahasiswa", dimana peneliti melakukan pengelompokan dari data penerimaan mahasiswa baru pada Universitas Narotama dengan melakukan *data mining*. Peneliti melakukan *clustering* dengan menggunakan algoritma K-means, serta aplikasi Weka 3.8 dalam melakukan penelitian. Hasil dari penelitian tersebut yaitu tiga buah *cluster*/kelompok, dimana dalam *cluster* pertama terdapat 1112 mahasiswa yang rata-rata berusia di antara 23 hingga 25 tahun, *cluster* kedua terdapat 825 mahasiswa dengan rata-rata usia 19 hingga 25 tahun, dan *cluster* ketiga terdapat 744 mahasiswa dengan rata-rata usia 25 tahun ke atas. Hasil yang diperoleh dipergunakan sebagai dasar dari tim pemasaran Universitas Narotama dalam membentuk strategi promosi mereka.

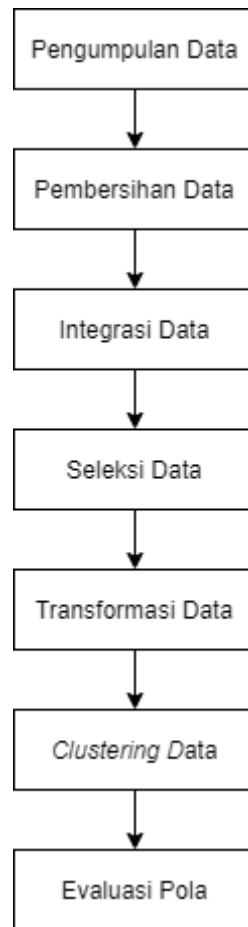
Jurnal penelitian yang dikembangkan oleh Sibarani & Omby (2018) berjudul "Algoritma K-Means Clustering Strategi Pemasaran Penerimaan Mahasiswa Baru Universitas Satya Negara Indonesia". Peneliti menggunakan metode *data mining*, dengan algoritma K-means untuk memperoleh pengelompokan mahasiswa berdasarkan dari data mahasiswa wisudawan Universitas Satya Negara Indonesia tahun 2016 dan 2017. Hasil dari penelitian digunakan bagian pemasaran dari Universitas Satya Negara Indonesia untuk membentuk strategi promosi yang tepat sasaran.

Jurnal penelitian yang dikembangkan oleh Chasanah & Widiyono (2017) berjudul "Penentuan Strategi Promosi Penerimaan Mahasiswa Baru Dengan Algoritma Clustering K-Means". Adapun yang dilakukan oleh peneliti yaitu *data mining* untuk menemukan *cluster-cluster* dari data pendaftaran mahasiswa tahun-tahun sebelumnya menggunakan algoritma K-means. Hasil dari penelitian digunakan untuk membantu dalam membentuk strategi promosi yang tepat oleh STMIK Widya Pratama Pekalongan.

Untuk penelitian ini, Penulis terinspirasi untuk melakukan *data mining* dalam Universitas XYZ, oleh jurnal penelitian dari Yunita (2018). Penulis juga terinspirasi untuk melakukan *data mining* dengan teknik *clustering*/pengelompokan data oleh jurnal penelitian dari Lestari (2019), serta melakukan *clustering* data menggunakan aplikasi Weka oleh jurnal penelitian dari Abriyanto & Damastuti (2019). Selain itu, Penulis mendasarkan ide untuk melakukan *clustering* data dengan menggunakan algoritma K-means oleh jurnal penelitian dari Sibarani & Omby (2018), dan jurnal penelitian dari Chasanah & Widiyono (2017).

Metodologi Penelitian

Adapun jenis penelitian yang akan dilakukan oleh Penulis yaitu penelitian kuantitatif. Penelitian kuantitatif merupakan jenis penelitian dimana data yang didapatkan untuk penelitian bersifat numerik dan angka-angka dari data tersebut dianalisa untuk memperoleh suatu kesimpulan (Nofiyani & Anggoro, 2017). Metode yang digunakan Penulis dapat penelitian ini yaitu dengan melakukan *data mining* menggunakan algoritma K-means. Alur dari penelitian yang dilakukan dapat dilihat pada Gambar 1.



Gambar 1. Alur Penelitian

Untuk melakukan *data mining*, Penulis harus terlebih dahulu mengumpulkan data yang dibutuhkan supaya dapat diproses nantinya. Data yang digunakan dalam penelitian tersebut merupakan data penerimaan mahasiswa Universitas XYZ. Data yang dikumpulkan akan langsung diperoleh dari pihak Universitas XYZ sendiri, dan akan digunakan dalam proses analisis data menggunakan algoritma K-means. Adapun atribut-atribut data yang diperlukan dalam data yang diperoleh yaitu jurusan mahasiswa di sekolah, wilayah asal, asal sekolah, jenis kelamin, prodi tujuan, pendidikan ayah, pekerjaan ayah, pendidikan ibu, pekerjaan ibu, gaji orang tua setiap bulan, sumber dana perkuliahan mahasiswa, serta jumlah saudara yang menjalankan studi di Universitas XYZ pula.

Setelah itu, data perlu dicek supaya dipastikan bahwa data dapat digunakan dan menghasilkan hasil yang bermanfaat. Apabila terdapat data yang jelek, maka perlu dibersihkan terlebih dahulu. Kemudian, apabila data yang diberikan berasal dari lebih dari satu sumber, maka semua data tersebut harus digabungkan menjadi satu terlebih dahulu. Setelah itu, perlu dilakukan seleksi hanya data tertentu yang sesuai untuk digunakan dalam proses *clustering*, dan perubahan data menjadi format atau bentuk yang dapat digunakan oleh aplikasi *data mining* dalam melakukan *clustering*. Kemudian, dilakukan *clustering* dari data tersebut. Untuk melakukan *data mining*, Penulis menggunakan aplikasi *data mining* yaitu Weka, yang dapat digunakan untuk melakukan berbagai macam *data mining* yang meliputi klasifikasi, *clustering*, dan asosiasi, dengan beraneka ragam algoritma yang dapat dipilih untuk digunakan. Akhirnya,

dalam tahap evaluasi pola, hasil *data mining* yang didapatkan dievaluasi lebih lanjut untuk menemukan pola-pola yang bermakna. Hasil *output* dari proses *clustering* yaitu model-model data yang berisi informasi dari masing-masing *cluster* yang telah dibentuk. Pola-pola yang ditemukan akan memperjelas informasi terkait seperti apakah anggota-anggota dari masing-masing *cluster*.

Hasil dan Pembahasan

Dalam data yang telah dikumpulkan, terdapat nilai-nilai yang tidak konsisten antara beberapa atribut yang terkait, serta terdapat nilai kosong pada data-data tertentu. Oleh karena itu, Penulis telah melakukan pembersihan data dengan mengubah nilai-nilai yang jelek supaya konsisten, serta memberikan nilai tertentu pada nilai yang kosong. Setelah itu, integrasi data dilakukan karena data diperoleh secara terpisah dari beberapa sumber, diikuti dengan tahap seleksi data, dimana Penulis memilih hanya variabel-variabel tertentu yang akan digunakan dalam penelitian ini. Kemudian, dilakukan transformasi data sesuai kebutuhan, supaya data dapat digunakan pada tahap berikutnya, yaitu tahap *clustering*, dimana data yang telah diolah akan diproses untuk mendapatkan hasil yang diinginkan.

Setelah semua tahap sebelumnya selesai dilakukan, data siap digunakan dalam tahap *clustering*, dengan menggunakan 10 buah *cluster*, serta empat variabel yang terpilih. Adapun empat buah variabel yang terpilih tersebut yaitu jurusan asal mahasiswa di sekolah sebelumnya, jenis kelamin, jumlah saudara mahasiswa yang menjalankan perkuliahan di universitas yang sama, dan asal wilayah. Hasil *clustering* dapat dilihat pada Tabel 1 dan Tabel 2.

Tabel 1. Hasil *Cluster* Pertama Hingga Kelima dari *Clustering* Data Mahasiswa

	Cluster 1 (910 anggota)	Cluster 2 (558 anggota)	Cluster 3 (304 anggota)	Cluster 4 (113 anggota)	Cluster 5 (1,082 anggota)
Jurusan asal	SMK Teknik	SMU/MA IPS	SMK Akuntansi	SMK Akuntansi	SMK Akuntansi
Jenis Kelamin	L	P	L	P	P
Jumlah saudara	0.0978	0.0502	0.1711	1.1593	0
Asal wilayah	A	C	A	A	A

Tabel 2. Hasil *Cluster* Keenam Hingga Kesepuluh dari *Clustering* Data Mahasiswa

	Cluster 6 (1,073 anggota)	Cluster 7 (821 anggota)	Cluster 8 (193 anggota)	Cluster 9 (939 anggota)	Cluster 10 (951 anggota)
Jurusan asal	SMU/MA IPA	SMU/MA IPA	SMK Akuntansi	SMU/MA IPS	SMU/MA IPS
Jenis Kelamin	P	L	P	L	P
Jumlah saudara	0.1342	0.1437	0.3005	0.2577	0.1767
Asal wilayah	A	A	K	A	A

Dari hasil *clustering* data seluruh mahasiswa (6,944 mahasiswa) menggunakan 10 *cluster* dan 4 variabel terpilih telah ditemukan 3 *cluster* terbesar yaitu sebagai berikut.

1. *Cluster* kelima yang terdapat 1,082 mahasiswa, dimana mahasiswanya lebih cenderung berasal dari jurusan SMK Akuntansi, berjenis kelamin perempuan, tidak memiliki saudara yang sedang kuliah di universitas yang sama, dan berasal dari daerah A.
2. *Cluster* keenam yang terdapat 1,073 mahasiswa, dengan mahasiswa yang rata-rata berasal dari jurusan SMU/MA IPA, berjenis kelamin perempuan, memiliki rata-rata jumlah saudara yang sedang kuliah di universitas yang sama 0.1342, dan berasal dari daerah A.
3. *Cluster* kesepuluh dengan 951 mahasiswa, dimana mahasiswanya rata-rata berasal dari jurusan SMU/MA IPS, berjenis kelamin perempuan, memiliki rata-rata jumlah saudara yang sedang kuliah di universitas yang sama yaitu 0.1767, dan berasal dari daerah A.

Berdasarkan hasil yang telah didapatkan di atas, rekomendasi dari Penulis yaitu mayoritas dari seluruh mahasiswa (4,527 dari 6,944 mahasiswa) masih berasal dari daerah A atau terindikasi berasal dari daerah A, oleh karena itu perlu untuk meluaskan penerimaan mahasiswa yang berasal dari luar daerah A. Selain itu, supaya dapat meningkatkan penerimaan mahasiswa baru, sebaiknya memprioritaskan penerimaan mahasiswa dari asal jurusan SMU/MA IPS yang paling mungkin berkuliah bersama dengan seorang atau lebih saudara di universitas yang sama, seperti dalam *cluster* di poin ketiga dari hasil ketiga *cluster* yang dijelaskan di atas.

Adapun batasan dari penelitian ini yaitu Penulis tidak dapat menyimpulkan data dengan sangat baik, dikarenakan persebaran data mahasiswa dalam setiap prodi kurang normal. Selain itu, terdapat data-data yang tidak lengkap sehingga hasil yang diperoleh belum mewakili semua kondisi yang ada.

Kesimpulan

Penerapan *data mining* dengan melakukan *clustering* menggunakan algoritma K-means dapat membantu Universitas XYZ dalam menentukan strategi promosi yang sesuai. Penulis melakukan *clustering* terhadap data penerimaan mahasiswa dalam Universitas XYZ, dan telah ditemukan 3 *cluster* yang mampu mempresentasikan mayoritas mahasiswa yang diterima dalam Universitas XYZ. Dari hasil evaluasi ketiga *cluster* tersebut, Penulis memberikan rekomendasi yaitu untuk meningkatkan penerimaan mahasiswa dari luar daerah A, serta mahasiswa yang berasal dari jurusan SMU/MA IPS. Saran dari Penulis untuk peneliti-peneliti selanjutnya di antaranya yaitu data-data yang akan diolah dan digunakan dalam proses *data mining* sebaiknya dilengkapi terlebih dahulu, supaya mampu memberikan hasil akhir yang lebih lengkap.

Daftar Pustaka

- Abriyanto, A., & Damastuti, N. (2019). Segmentasi Mahasiswa Dengan Unsupervised Algoritma Guna Membangun Strategi Marketing Penerimaan Mahasiswa. *Insand Comtech: Information Science and Computer Technology Journal*, 4(2), 10–18.
- Akbar, R., Yuliani, E., Mawaddah, Q., & Ardhana, F. (2017). Analisis Data Penjualan Perusahaan Detergen XYZ dengan Aplikasi Zoho Reporting Menggunakan Metode OLAP (Online Analytical Processing). *Jurnal Edukasi Dan Penelitian Informatika (JEPIN)*, 3(1), 71–75.
- Andreanus, J., & Kurniawan, A. (2017). Sejarah, Teori Dasar dan Penerapan Reinforcement Learning: Sebuah Tinjauan Pustaka. *Jurnal Telematika*, 12(2), 113–118.
- Arifin, M. (2020). Implementasi Data Mining Pada Prediksi Pemesanan Menggunakan Algoritma Apriori (Studi Kasus: Kimia Farma). *Pelita Informatika: Informasi Dan Informatika*, 8(3), 353–356.
- Chasanah, T. T., & Widiyono. (2017). Penentuan Strategi Promosi Penerimaan Mahasiswa Baru Dengan Algoritma Clustering K-Means. *Jurnal IC-Tech*, 12(1), 39–44.

- Gustientiedina, Adiya, M. H., & Desnelita, Y. (2019). Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan. *Jurnal Nasional Teknologi Dan Sistem Informasi*, 5(1), 17–24.
- Lestari, W. (2019). Clustering Data Mahasiswa Menggunakan Algoritma K-Means Untuk Menunjang Strategi Promosi (Studi Kasus: STMIK Bina Bangsa Kendari). *Jurnal Sistem Informasi Dan Sistem Komputer*, 4(2), 35–48.
- Mardi, Y. (2017). Data Mining: Klasifikasi Menggunakan Algoritma C4.5. *Jurnal Edik Informatika*, 2(2), 213–219.
- Nofiyani, & Anggoro, D. (2017). Pengembangan Sistem Penunjang Keputusan Berbasis Web Untuk Menentukan Best Customer dengan Model RFM (Recency, Frequency dan Monetary), Metode Comparative Performance Index Dan Algoritma K-Means. *Jurnal Sistem Komputer*, 7(1), 13–20.
- Sibarani, R., & Omby. (2018). Algoritma K-Means Clustering Strategi Pemasaran Penerimaan Mahasiswa Baru Universitas Satya Negara Indonesia. *Jurnal Algoritma, Logika Dan Komputasi*, 1(2), 44–50.
- Syahril, M., Erwanyah, K., & Yetri, M. (2020). Penerapan Data Mining Untuk Menentukan Pola Penjualan Peralatan Sekolah Pada Brand Wigglo Dengan Menggunakan Algoritma Apriori. *Jurnal Teknologi Sistem Informasi Dan Sistem Komputer TGD*, 3(1), 118–136.
- Yunita, F. (2018). Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Pada Penerimaan Mahasiswa Baru. *SISTEMASI: Jurnal Sistem Informasi*, 7(3), 238–249.