Jitu: Jurnal Informatika Utama
Vol. 3, No. 1 Mei 2025
e-ISSN 2988-7631
Hal: 112-126

PENERAPAN DATA MINING UNTUK MENENTUKAN LOKASI PROMOSI SEKOLAH DENGAN METODE K-MEANS CLUSTERING (STUDI KASUS: SMP ISLAM AL SYUKRO UNIVERSAL)

Ikhwan Fauzi
Fakultas Komputer Sistem Informasi, Universitas Pamulang, Tangerang Selatan, Banten, Indonesia
Email: dosen03150@unpam.ac.id

ABSTRAK

Setiap sekolah menginginkan peningkatan dalam penerimaan siswa baru pada setiap tahunnya. Sehingga setiap sekolah memiliki strategi masing-masing dalam menarik minat para calon siswanya, serta menentukan daerah mana saja yang menjadi lokasi potensial dalam melakukan promosi. Penelitian ini bermaksud untuk menentukan lokasi atau daerah mana saja yang berpotensi untuk mendatangkan siswa baru di SMP Islam Al Syukro Universal. Dengan menerepkan data mining dengan metode *Clustering* mengelompokkan item objek penelitian berdasarkan kemiripan sifatnya, sehingga akan diketahui daerah mana saja yang berpotensi untuk mendatangkan siswa baru. Penentuan lokasi promosi penerimaan siswa baru dengan metode data mining akan memberikan dampak yang bagus dan terarah dalam melakukan promosi, sehingga dapat meningkatkan jumlah siswa baru setiap tahunnya.
Kata Kunci: Data Mining, Metode Clustering, Promosi

ABSTRACT

Every school wants an increase in new student admissions every year. So that each school has its own strategy in attracting the interest of prospective students, as well as determining which areas are potential locations for promotion. This study intends to determine which locations or areas have the potential to bring in new students at Al Syukro Universal Islamic Junior High School. By applying data mining with the Clustering method, the research object items are grouped based on their similarity, so it will be known which areas have the potential to bring in new students. Determining the location of the promotion of new student admissions with the data mining method will have a good and directed impact in carrying out promotions, so that it can increase the number of new students every year.
Keywords: *Data Mining, Clustering Method, Promotion*

PENDAHULUAN

Data mining merupakan proses untuk mendapatkan informasi dari basis data yang besar dengan cara penggalian data dengan memanfaatkan kumpulan data. Dengan kemajuan teknologi informasi sekarang sudah semakin berkembang pesat dan hampir mencangkup di segala bidang. Kemajuan teknologi tersebut menghasilkan data yang besar dan banyak mulai dari bidang industri, ekonomi, ilmu dan teknologi serta berbagai bidang lainnya. Penerapan teknologi informasi dalam dunia pendidikan juga dapat menghasilkan data yang berlimpah mengenai data dari siswa yang mengikuti pendidikan.

Misalnya dalam proses penerimaan siswa baru sebuah sekolah akan menghasilkan data baru berupa profil dari siswa baru. Hal seperti ini akan terjadi

Creative Commons Attribution 4.0 International License

secara berulang pada sebuah perguruan tinggi. Jika dilakukan pengolahan data pada sumber data tersebut maka dapat diketahui berbagai informasi yang bermanfaat dalam membantu menentukan strategi promosi penerimaan siswa baru tahun berikutnya.

Seiring perkembangan teknologi dan informasi yang semakin pesat banyak bermunculan sekolah-sekolah dengan menghadirkan konsep-konsep pendidikan yang baru. Hal tersebut membuat sekolah-sekolah diberbagai tempat di indonesia berlomba-lomba untuk membuat sekolah mereka agar banyak diminati oeh pelajar. Berbagai macam strategi atau cara mereka lakukan untuk menarik minat peserta didik agar bersekolah ditempat mereka. Dengan adannya persaingan ini maka pihak sekolah memunculkan strategi pemasaran sekolah atau yang lebih dikenal dengan bahasa asing yaitu "Marketing". Istilah marketing ini tidak hanya dikenal dalam dunia bisnis, industri, dan perdagangan akan tetapi meluas kepada dunia pendidikan yakni sekolah.

Pada penelitian ini penulis memilih SMP Islam Al Syukro Universal sebagai objek yang diteliti. Atribut yang digunakan diantaranya alamat siswa, sekolah asal yang diambil serta yang terakhir jarak lokasi rumah ke sekolah. Pada daerah asal penulis akan lebih menyepesifikasikan kelurahan. hal ini dilakukan dengan suatu pertimbangan supaya dalam promosi terarah tidak melebar kemana-mana dan sesuai target yang sudah ditentukan

SMP Islam Al Syukro Universal merupakan salah satu sekolah swasta yang terus menerus melakukan peningkatan baik dari segi kualitas pendidikan maupun kualitas pembangunan sekolah.

Karena siswa SMP Islam Al Syukro Universal berasal dari sekitar daerah Ciputat maupun luar daerah Ciputat maka dibutuhkan strategi khusus oleh marketing melakukan pemasaran untuk mencari calon siswa-siswi agar promosi yang dilakukan lebih efektif dan efisien. Selama ini untuk penyebaran pamflet masih disebarkan disembarang tempat dan promosi tidak terorganisir maka sangat tidak efisien untuk pemasaran. Untuk menjawab semua pertanyaan tersebut akan diadakan sebuah penelitian yang dilakukan dengan cara mengolah data-data siswa yang masih aktif maupun yang telah lulus seperti nama siswa, alamat siswa, tingkat sekolah yang diambil serta yang terakhir nilai akademik siswa yang merupakan salah satu faktor literatur penting dalam dunia pendidikan. Data-data yang berhasil didapatkan tadi kemudian diolah untuk mengetahui pola dari data-data tersebut sehingga kita dapat mengambil informasi-informasi yang tersembunyi dari data-data tersebut.

Berikut merupakan beberapa penelitian yang pernah dilakukan terkait Penerapan Data Mining Untuk Menentukan Lokasi Promosi Sekolah dengan Metode K-Means Clustering. Yaitu sebagai berikut:

Menurut penelitian yang dilakukan oleh Felix Toknady Kesuma, Feryanto, Jurnal Sistem Informasi Ilmu Komputer Prima (JUSIKOM PRIMA) Vol. 2 No. 2, Maret 2019 e-ISSN :2580:2879 dengan judul "Penerapan Data Mining Untuk Menentukan Penjualan Sparepart Toyota Dengan Metode K-Means Clustering" Berdasarkan hasil pengelompokan dari seluruh data menggunakan metode k-means clustering, didapatkan hasil akhir pengelompokan hingga iterasi ke-5, dimana titik pusat tidak lagi berubah dan tidak ada data yang berpindah antar cluster.

Creative Commons Attribution 4.0 International License

Penelitian yang dilakukan oleh Ilka Zufria, JISTech, Vol.3, No.1, Januari-Juni 2018 ISSN: 2528-5718 dengan judul "Penentuan Potensi Lokasi Promosi Calon Mahasiswa Baru Perguruan Tinggi Swasta Berbasis AHP (*Analytical Hierarchy Process*)" Berdasarkan penelitian yang telah dilakukan baik yang dihitung secara manual maupun dengan Software Expert Choice untuk mengimplementasikan metode AHP dalam penentuan potensi promosi pada siswa-siswi SMA atau setingkatnya untuk masuk ke sebuah PTS dapat disimpulkan adalah Model kriteria dalam menentukan potensi pemasaran pada 5 Sekolah untuk masuk ke sebuah PTS ditentukan 4 kriteria yaitu : jarak lokasi sekolah terhadap kampus, prosentase minat murid SMA untuk masuk kampus, kerjasama(MOU) yang pernah dilakukan terhadap Sekolah, dan jumlah murid SMA atau setingkatnya yang masuk pada sebuah PTS pada tahun sebelumnya. Adapun hasil perhitungan untuk masing-masing kriteria dihasilkan 57%, 25%, 10% dan 8%.

Penelitian yang dilakukan oleh Widodo, Dina Wahyuni, Jurnal PINTER Vol. 1 No.2 Desember 2017 e-ISSN: 2597- 4475 dengan judul "Implementasi Algoritma K-Means Clustering Untuk Mengetahui Bidang Skripsi Mahasiswa Multimediapendidikan Teknik Informatika Dan Komputer Universitas Negeri Jakarta" Berdasarkan hasil penelititian yang telah dilakukan dapat disimpulkan yaitu Program bantu yang dikembangkan dari alur kerja implementasi algoritma K-Means dapat digunakan untuk mengetahui pengelompokan bidang skripsi mahasiswa multimedia PTIK UNJ angkatan 2010 yaitu bidang video dan animasi serta bidang aplikasi pengembangan perangkat lunak. Penentuan centroid awal pada proses clustering dapat mempengaruhi proses iterasi. Semakin dekat nilai centroid awal dengan nilai mean data, maka proses iterasi yang dihasilkan sedikit. Sebaliknya jika nilai centroid awal jauh dari nilai mean data, maka iterasi yang dihasilkan lebih banyak.

Penelitian yang dilakukan oleh Fitri Yunita, Jurnal SISTEMASI, Volume 7, Nomor 3 September 2018 : 238 – 249 E-ISSN:2540-9719 dengan judul "Penerapan Data Mining Menggunkan Algoritma *K-Means Clustring* Pada Penerimaan Mahasiswa Baru (Studi Kasus : Universitas Islam Indragiri)" Setelah banyak tahapan dilakukan dalam penerapan Algoritma *K-Means Clustring* ada beberapa kesimpulan yaitu , Penentuan centroid (titik pusat) pada tahap awal Algoritma *K-Means* sangat berpengaruh pada hasil cluster seperti pada hasil Pengujian yang dilakukan dengan menggunakan 516 dataset dengan centroid yang berbeda menghasilkan hasil cluster yang berbeda juga. Setelah dilakukan pengolompokan data penerimaan mahasiswa menggunakan metode k-means Clustring terbentuk tiga cluster yaitu cluser satu dengan jumlah 195 items, cluster dua 271 dengan jumlah 271 items, dan cluster tiga dengan jumlah 50 items. Strategi promosi bagi calon mahasiswa baru nantinya akan mengikuti cluster yang terbentuk bardasarkan program studi yang paling banyak diminati di masing-masing sekolah.

Penelitian yang dilakukan oleh Ramdani Budiman, Rudianto, rnal ProTekInfo Vol. 6 No 1 | September 2019 ISSN : 2406-7741 dengan judul "Penerapan Data Mining Untuk Menentukan Lokasi Promosi Penerimaan Mahasiswa Baru Pada Universitas Banten Jaya (Metode *K-Means Clustering*)" Berdasarkan hasil Analisa yang di dapat dengan metode clustering k-means dalam menentukan pengelompokan wilayah instansi pendaftar yang berada di daerah kota serang, dapat di simpulkan bahwa

Creative Commons Attribution 4.0 International License

Setelah dilakukan pengelompokan data calon mahasiswa melalui persebaran berdasarkan wilayah menggunakan K-Means clustering terbentuk menjadi tiga Cluster.Cluster pertama kategori tinggi, cluster kedua kategori rendah dan cluster tiga kategori sedang. Strategi promosi bagi calon mahasiswa baru yang tepat sasaran untuk setiap wilayah berdasarkan cluster terbentuk. Setelah terbentuk cluster dengan metode kmeans maka mengirim tim yang sesuai dengan program studi yang paling banyak diminati dan melakukan promosi berdasarkan peminatan calon mahasiswa dengan melakukan penyelarasan. Setelah didapatkan hasil dari pengelompokanberdasarkan cluster maka sistem akanmenampilkan hasil pengelompokan wilayahcalon mahasiswa dalam bentuk dashboard.

Penelitian yang dilakukan oleh Michael Rusli, Satria Perdana Arifin, Anggy Trisnadoli, Jurnal Komputer Terapan Vol. 3, No. 1, Mei 2017, 11-18 dengan judul "Pengembangan Sistem Pendukung Keputusan untuk Penentuan Lokasi Promosi Penerimaan Mahasiswa Baru" Sistem yang dibangun telah lulus uji User Acceptance Test yang dilakukan sebanyak dua kali. Dari 18 butir pengujian bahwa semua butir pengujian sistem sudah berjalan dengan benar dan dapat diartikan sistem dapat diterima oleh Staf Marketing maupun Tim PMB dan Pimpinan. Sistem dapat mendukung sebuah keputusan mengenai lokasi promosi yang memiliki potensi untuk mendapatkan mahasiswa baru dengan lebih cepat, teliti, dan tepat sasaran karena bersifat dinamis serta mampu membantu menghitung efisiensi biaya marketing setelah dilakukan pengujian oleh Ketua PMB dan Koordiantor Sosialisasi. Sistem ini telah menerapkan metode AHP untuk penyeleksian pada tahap pertama dan telah menerapkan metode TOPSIS untuk penyeleksian pada tahap kedua. Selanjutnya dilakukan uji akurasi dengan membandingkan hasil manual dengan hasil sistem dan diperoleh hasil perbandingan sebesar 100%. Untuk pengembangan penelitian dimasa mendatang, sistem ini dapat dikembangkan dengan menambahkan altenatif sekolah-sekolah yang dengan cakupan yang lebih luas lagi, seperti menggunakan alternatif kecamatan, dan provinsi yang lebih banyak sehingga mendapatkan data yang lebih besar. Sistem ini juga dapat dikembangkan dengan menambahkan fitur pengambilan data altenatif dan kriteria yang lebih flexible, seperti menggunakan masukan dari file tertentu, masukan manual satuan maupun masukan dalam jumlah besar dengan menggunakan database yang terpisah dari sistem yang berjalan.

Penelitian yang dilakukan oleh Elvira Asril, Fana Wiza, Yogi Yunefri, urnal Teknologi Informasi & Komunikasi *Digital Zone*, Volume 6, Nomor 2, November 2015: 24-32 dengan judul "Analisis Data Lulusan dengan Data Mining untuk Mendukung Strategi Promosi Universitas Lancang Kuning" Setelah dilakukan pengelompokan data mahasiswa melalui persebaran wilayah berdasarkan potensi akademik menggunakan K-Means clustering terbentuk empat cluster. Strategi promosi bagi calon mahasiswa baru yang tepat sasaran untuk setiap wilayah berdasarkan cluster yang terbentuk adalah dengan mengirim tim promosi berdasarkan potensi akademik mahasiswa dengan melihat rata-rata IPK pada setiap cluster.

METODE

Metode pengolahan data yang digunakan dalam penelitian ini sering disebut sebagai data mining. Analisis yang diterapkan adalah metode K-Means Clustering,

Creative Commons Attribution 4.0 International License

yaitu suatu metode yang mengelompokkan data berdasarkan karakteristik yang sama ke dalam satu cluster, sementara data dengan karakteristik berbeda akan dikelompokkan ke cluster lain yang sesuai. Hasil dari pengolahan data tersebut sangat membantu pihak sekolah dalam memperoleh informasi sebaran siswa berdasarkan data masuk setiap angkatan. Dengan demikian, dapat diketahui wilayah mana yang menjadi prioritas untuk promosi calon siswa baru di masa mendatang. Algoritma K-Means memiliki beberapa kelebihan, seperti kemudahan dalam implementasi dan eksekusi, waktu proses yang relatif cepat, kemampuan adaptasi yang baik, serta penggunaannya yang umum dalam berbagai kasus. Dengan pengelompokkan data yang telah dilakukan, diharapkan pihak marketing dapat melakukan strategi pemasaran yang lebih terarah dan efektif sesuai dengan target yang telah ditentukan untuk mendapatkan calon siswa baru.

Berdasarkan latar belakang tersebut, permasalahan utama yang diidentifikasi adalah adanya persaingan ketat akibat banyaknya sekolah swasta yang bermunculan sehingga dibutuhkan strategi promosi yang tepat agar tidak terjadi penurunan jumlah penerimaan siswa baru. Selain itu, pengelompokkan data yang melibatkan siswa aktif dan alumni membutuhkan metode pengolahan data yang tepat untuk memperoleh hasil yang optimal. Untuk membatasi ruang lingkup penelitian agar tetap fokus dan tidak menyimpang dari masalah utama, penelitian ini hanya berfokus pada penentuan strategi promosi di SMP Islam Al Syukro Universal melalui implementasi metode Data Mining dengan menggunakan algoritma K-Means Clustering. Pengujian dilakukan menggunakan tool WEKA dengan data siswa yang masih aktif maupun yang sudah lulus.

Rumusan masalah yang dihadapi dalam penelitian ini meliputi bagaimana cara menganalisis penentuan lokasi promosi yang tepat dengan menerapkan metode Data Mining menggunakan algoritma K-Means Clustering, serta bagaimana cara mengelompokkan data siswa berdasarkan atribut yang digunakan untuk menentukan lokasi promosi di setiap daerah. Tujuan penelitian ini adalah menerapkan metode Data Mining dengan algoritma K-Means Clustering ke dalam sebuah aplikasi untuk menentukan wilayah promosi yang tepat bagi SMP Islam Al Syukro Universal serta mengetahui persebaran wilayah berdasarkan atribut siswa yang digunakan dalam analisis. Manfaat yang diharapkan dari penelitian ini adalah kemampuan menganalisis lokasi promosi yang efektif bagi SMP Islam Al Syukro Universal dengan menggunakan metode yang diterapkan dalam penelitian ini. Pengelompokkan data siswa dalam setiap cluster yang terbentuk dapat dijadikan sebagai dasar pengambilan keputusan untuk menentukan wilayah promosi yang tepat sasaran dalam penerimaan siswa baru.

HASIL PENELITIAN

Menurut Han dan Kamber (2011, p36), data mining adalah proses menemukan pola yang menarik dan pengetahuan dari data yang berjumlah besar. Sedangkan menurut Linoff dan Berry (2011, p7) Data mining adalah suatu pencarian dan analisa dari jumlah data yang sangat besar dan bertujuan untuk mencari arti dari pola dan aturan. Lalu menurut Connolly dan Begg (2010), Data mining adalah suatu proses ekstraksi atau penggalian data yang belum diketahui sebelumnya, namun dapat dipahami dan berguna dari *database* yang besar serta digunakan untuk membuat

Creative Commons Attribution 4.0 International License

suatu keputusan bisnis yang sangat penting. Dan menurut Vercellis (2009, p77), *Data mining* adalah aktivitas yang menggambarkan sebuah proses analisis yang terjadi secara iteratif pada database yang besar, dengan tujuan mengekstrak informasi dan *knowledge* yang akurat dan berpotensial berguna untuk *knowledge workers* yang berhubungan dengan pengambilan keputusan dan pemecahan masalah.

Algoritma K-Means Clustering

K-Means merupakan salah satu algoritma clustering. Tujuan algoritma ini yaitu untuk membagi data menjadi beberapa kelompok. Algoritma ini menerima masukan berupa data tanpa label kelas. Hal ini berbeda dengan supervised learning yang menerima masukan berupa vektor (x1, y1), (x2, y2), …, (xi, yi), di mana xi merupakan data dari suatu data pelatihan dan yi merupakan label kelas untuk xi. Pada algoritma pembelajaran ini, komputer mengelompokkan sendiri data-data yang menjadi masukannya tanpa mengetahui terlebih dulu target kelasnya. Pembelajaran ini termasuk unsupervised learning. Masukan yang diterima adalah data atau objek dan k buah kelompok (cluster) yang diinginkan. Algoritma ini akan mengelompokkan data atau objek ke dalam k buah kelompok tersebut. Pada setiap cluster terdapat titik pusat (centroid) yang merepresentasikan cluster tersebut.

K-Means ditemukan oleh beberapa orang yaitu Lloyd (1957, 1982), Forgey (1965), Friedman and Rubin (1967), and McQueen (1967). Ide dari clustering pertama kali ditemukan oleh Lloyd pada tahun 1957, namun hal tersebut baru dipublikasi pada tahun 1982. Pada tahun 1965, Forgey juga mempublikasi teknik yang sama sehingga terkadang dikenal sebagai Lloyd-Forgy pada beberapa sumber.

K-Means merupakan metode klasterisasi yang paling terkenal dan banyak digunakan di berbagai bidang karena sederhana, mudah diimplementasikan, memiliki kemampuan untuk mengklaster data yang besar, mampu menangani dataoutlier, dan kompleksitas waktunya linear O(nKT) dengan n adalah jumlah dokumen, K adalah jumlah kluster, dan T adalah jumlah iterasi. K-Means merupakan metode pengklasteran secara partitioning yang memisahkan data ke dalam kelompok yang berbeda. Dengan partitioning secara iteratif, K-Means mampu meminimalkan rata-rata jarak setiap data ke klasternya. Metode ini dikembangkan oleh Mac Queen pada tahun 1967.

Dasar algoritma K-Means adalah sebagai berikut:

1. Tentukan nilai k sebagai jumlah klaster yang ingin dibentuk
2. Bangkitkan k centroid (titik pusat klaster) awal secara random
3. Hitung jarak setiapdata ke masing-masing centroid menggunakan Rumus korelasi antar dua objek yaitu Euclidean Distance dan kesamaan Cosine
4. Kelompokkan setiap data berdasarkan jarak terdekat antara data dengan centroidnya
5. Tentukan posisi centroid baru (Ck) dengan cara menghitung nilai rata-rata dari data-data yang ada pada centroid yang sama.

$$C_k = \left(\frac{1}{n_k}\right) \sum d_i$$

Creative Commons Attribution 4.0 International License

Dimana *nk* adalah jumlah dokumen dalam cluster k dan *di* adalah dokumen dalam cluster k. Kembali ke langkah 3 jika posisi centroid baru dengan centroid lama tidak sama.

Adapun karakteristik dari algoritma K-Means salah satunya adalah sangat sensitif dalam penentuan titik pusat awal klaster karena K-Means membangkitkan titik pusat klaster awal secara random. Pada saat pembangkitan awal titik pusat yang random tersebut mendekati solusi akhir pusat klaster, K-Means mempunyai posibilitas yang tinggi untuk menemukan titik pusat klaster yang tepat. Sebaliknya, jika awal titik pusat tersebut jauh dari solusi akhir pusat klaster, maka besar kemungkinan ini menyebabkan hasil pengklasteran yang tidak tepat. Akibatnya K-Means tidak menjamin hasil pengklasteran yang unik. Inilah yang menyebabkan metode K-Means sulit untuk mencapai optimum global, akan tetapi hanya minimum lokal. Selain itu, algoritma K-Means hanya bisa digunakan untuk data yang atributnya bernilai numerik.

***Waikato Environment of Knowledge* (WEKA)**

Sigit Adinugroho dan Yuita Arum Sari (2018:8) mengemukan bahwa WEKA merupakan software terintegrasi yang berisi implementasi dari metode-metode data mining. WEKA dikembangkan oleh Universitas Waikato, Selandia Baru menggunakan bahasa pemrograman Java. Oleh karena itu, Weka merupakan singkatan dari *Waikato Environment for Knowledge Analysis*. Dengan mengadopsi konsep opensource software, menjadikan Weka dapat digunakan dan dimodifikasi siapapun secara gratis.

WEKA memiliki keunggulan jika dibandingkan dengan perangkat lunak data mining lainnya. Penggunaan WEKA murah karena aplikasi tersebut berlilensi GNU *General Public License*, yang artinya dapat digunakan secara gratis. Penggunaan bahasa Java dalam pengembangan WEKA menyebakan WEKA dapat diinstal pada hampir semua sistem operasi modern, sepanjang sistem operasi tersebut mendukung *Java Virtual Machine*.

WEKA menyediakan tiga antarmuka grafis untuk mengolah data. Tampilan pertama adalah explorer yang menyediakan akses ke semua fungsi-fungsi pada Weka melalui menu-menu yang urut dan mudah digunakan. Fungsi yang dapat digunakan melalui explorer adalah:

1. *Preprocessing*
   Merupakan panel untuk memilih data yang akan diproses. Data dapat berbentuk berkas ARFF, database, maupun data buatan yang dibangkitkan dengan pola tertentu.
2. *Classifier*
   Classifier merupakan panel untuk menggunakan algoritma klasifikasi atau regresi terhadap data yang telah dimasukkan di panel preprocessing. Terdapat bermacam-macam algoritma klasifikasi dan regresi yang telah disediakan dan siap digunakan. Terdapat pula sejumlah metode evaluasi yang dapat digunakan untuk mengevaluasi hasil klasifikasi.
3. *Cluster*

Creative Commons Attribution 4.0 International License

Merupakan panel yang memberikan akses terhadap berbagai metode clustering yang disediakan Weka. Panel ini memiliki beberapa opsi pembagian data menjadi data latih dan data uji untuk keperluan pengujian hasil clustering.

4. *Associate*
   Menyediakan algoritma untuk association rule yang berguna untuk mengenali relasi antar item pada data.
5. *Select attribute*
   Menyediakan berbagai algoritma untuk memilih atribut atau variabel yang paling relevan untuk suatu permasalahan

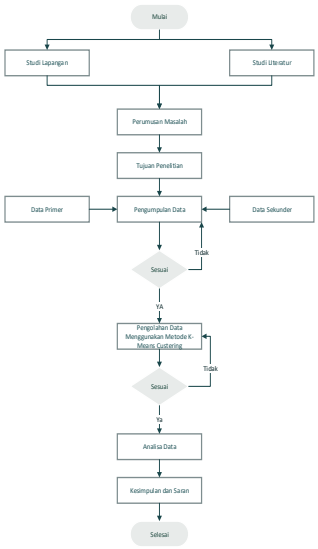
SSE *(Sum of Square Error)*

Sum Squared Error (SSE) merupakan salah satu metode untuk mengukur kesalahan atau error dalam suatu penelitian. Singkatnya SSE merupakan jumlah error yang dikuadratkan. Algoritma K-means mengelompokkan sekumpulan data sehingga fungsi SSE dapat diminimalkan. Error yang didapatkan biasanya diukur dengan menggunakan Euclidean Distance antara titik data atau objek dengan titik mean pada cluster. Sehingga SSE yang dihitung merupakan jumlah dari selisih titik data tiap objek dengan nilai mean pada clusternya yang dikuadratkan. Misalkan μk adalah mean dari kelompok k (Ck) maka *Sum Squared Error* untuk semua cluster dapat dinyatakan sebagai berikut:

$$SSE = \sum_{k=1}^{K} \sum_{Xi \in Ck} ||X_i - \mu_k||^2$$

Dimana μk adalah mean dari cluster k dan Xi adalah nilai objek ke-i pada cluster k.

1.1. Perancangan Penelitian



Gambar 3.1 Rancangan Penelitian

1.2. Pre-processing Data

Creative Commons Attribution 4.0 International License

Tahap pre-processing data adalah tahap dimana data yang sudah didapatkan, dipilah, dan dipisahkan agar mendapatkan data yang dibutuhkan untuk proses selanjutnya. Tahapan ini mempunyai beberapa proses dimana setiap proses tersebut saling berhubungan satu sama lainnya. Proses dalam tahapan preprocessing adalah sebagai berikut:

1. *Data Reduction*
   Data reduction adalah proses untuk mereduksi atau mengurangi dimensi, atribut, ataupun sejumlah data yang tidak dibutuhkan dalam suatu file data. Datareduction sangat berguna untuk mendapatkan atribut dan sejumlah data yang akan digunakan di dalam penelitian ini.
2. *Data Cleaning*
   Data cleaning adalah proses dalam tahapan preprocessing untuk mengisi data kosong atau blank apabila memungkinkan, duplikasi data, memperbaiki data yang tidak sesuai dengan ketentuan atau salah ketik seperti kurang huruf dan kelebihan huruf, mengubah dan memodifikasi data agar data yang akan diolah adalah data yang konsisten, mengatur data yang kurang rapi dalam penulisan huruf besar dan kecil, dan mengganti format penulisan angka dan huruf sesuai dengan yang dibutuhkan.
3. *Data Integration*
   Data integration adalah suatu proses untuk menggabungkan atau mengintegrasikan data dari beberapa file sumber. Data integration hanya dilakukan jika data berasal dari tempat yang berbeda-beda (sumber data tidak hanya dari 1 tempat). Langkah yang dilakukan antara lain mengintegrasikan skema, mengidentifikasi masalah entitas, dan mendeteksi sekaligus menyelesaikan konflik pada nilai data.
4. *Data Transformation*
   Karena metode *K-Means Clustering* adalah metode yang bisa dilakukan apabila data yang dipakai adalah data berupa angka, maka proses transformation ini sangat dibutuhkan. Proses transformation adalah tahap untuk mengubah data atribut yang selain angka ke dalam nilai angka agar data tersebut dapat diolah menggunakan algoritma *K- Means Clustering*.

K-Means Clustering

Setelah data melalui tahap pre-processing dan tahap transformation maka data sudah siap untuk diolah menggunakan *metode K-Means Clustering*. Metode Clustering yang digunakan adalah non-hierarchical Clustering dengan menentukan jumlah K cluster terlebih dahulu, dimana peneliti akan mencoba 2 cluster yaitu K=3 dan K=5, dan akan menentukan jumlah cluster mana yang terbaik. Machine mining yang digunakan dalam proses data mining dengan Algoritma K-Means Clustering adalah tool WEKA.

Penerapan Algroritma *K-means Clustering*

Metode *Clustering* yang digunakan adalah *non-hierarchical Clustering* dengan menentukan jumlah K cluster terlebih dahulu, dimana peneliti akan mentukan 2 jumlah cluster yaitu K=3 dan K=5, dan akan menentukan jumlah cluster mana yang terbaik. Machine mining yang digunakan dalam proses data mining dengan

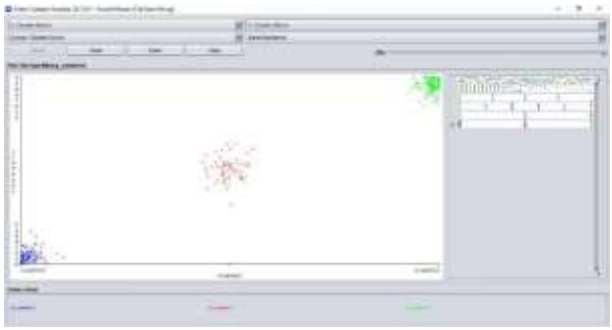
Creative Commons Attribution 4.0 International License

Algoritma K-Means Clustering adalah tool WEKA. Berikut ini adalah hasil clustering kombinasi K=3 dan K5:

```
kMeans
======

Number of iterations: 13
Within cluster sum of squared errors: 38.21077166311945

Initial starting points (random):

Cluster 0: 6,2,3,1
Cluster 1: 6,1,3,1
Cluster 2: 2,1,1,1

Missing values globally replaced with mean/mode

Final cluster centroids:
                        Cluster#
Attribute    Full Data          0          1          2
                (297.0)     (95.0)     (62.0)    (140.0)
=========================================================
Alamat           7.0707    12.9789     5.0806     3.9429
Sekolah          1.7811     1.8316     1.7742       1.75
Pekerjaan        1.9697     1.6632     3.9355     1.3071
Jarak            1.3939     2.1368     1.1452          1
```

Gambar 4.1 Hasil clustering K=3 dengan Tool WEKA

Pada gambar 4.1 menunjukan hasil clustering data siswa SMP Islam Al Syukro Universal dengan jumlah clustering K=3 melakukan iterasi sebanyak 13 kali dan menggasilkan *Sum of Squared Error* (SSE) sebesar 38,21077166311945 dengan jumlah anggota pada setiap cluster adalah cluster0 95 data, cluster1 62 data dan cluster2 140 data.



Gambar 4.2 Hasil Grafik K=3 dengan Tool WEKA

Berdasarkan hasil proses clustering menggunakan K-Means maka didapatkan sebuah informasi untuk penentuan promosi sebagai berikut:

1. Cluster 1 (C0)
   Cluster 1 dengan jumlah 95 data siswa didominasi Sekolah Islam (SDI), Untuk alamat didominasi Pamulang Timur dan Pondok Benda dengan Jarak di atas 5 KM dan pekerjaan rata-rata adalah Karyawan Swasta.
2. Cluster 2 (C1)
   Cluster 2 dengan jumlah 62 data siswa didominasi Sekolah Islam (SDI), Untuk alamat didominasi Cipayung dan Kedaung dengan jarak di bawah 5 KM dan pekerjaan rata-rata adalah Pegawai Negeri.
3. Cluster 3 (C2)

Creative Commons Attribution 4.0 International License

Cluster 3 dengan jumlah 140 data siswa didominasi Sekolah Islam (SDI), Untuk alamat didominasi Kedaung, Cipayung dan Ciputat dengan jarak di bawah 5 KM dan pekerjaan rata-rata adalah Karyawan Swasta.

```
kMeans
======

Number of iterations: 9
Within cluster sum of squared errors: 28.153419247762702

Initial starting points (random):

Cluster 0: 6,2,3,1
Cluster 1: 6,1,3,1
Cluster 2: 2,1,1,1
Cluster 3: 7,2,1,2
Cluster 4: 4,1,2,1

Missing values globally replaced with mean/mode

Final cluster centroids:
                          Cluster#
Attribute    Full Data          0          1          2          3          4
               (297.0)     (46.0)     (36.0)     (76.0)     (83.0)     (56.0)
==============================================================================
Alamat          7.0707     5.3261     5.7222     4.0921    13.4096     4.0179
Sekolah         1.7811     3.4565       1.75     1.3158     1.5663      1.375
Pekerjaan       1.9697     1.4783     4.6389          1     1.6988      2.375
Jarak           1.3939     1.2174     1.2778          1     2.1687          1
```

Gambar 4.3 Hasil clustering K=5 dengan Tool WEKA

Pada gambar 4.3 menunjukan hasil clustering data siswa SMP Islam Al Syukro Universal dengan jumlah clustering K=5 melakukan iterasi sebanyak 9 kali dan menggasilkan *Sum of Squared Error* (SSE) sebesar 28,153419247762702 dengan jumlah anggota pada setiap cluster adalah cluster0 46 data, cluster1 36 data, cluster2 76 data, cluster3 83 data dan cluster4 56 data.

Gambar 4.4 Hasil Grafik K=5 dengan Tool WEKA

Berdasarkan hasil proses clustering menggunakan K-Means maka didapatkan sebuah informasi untuk penentuan promosi sebagai berikut:

1. Cluster 1 (C0)
   Cluster 1 dengan jumlah 46 data siswa didominasi Sekolah Islam (SDI), Untuk alamat didominasi Cipayung, Pamulang Barat dan Pamulang Timur dengan Jarak di bawah 5 KM dan pekerjaan rata-rata adalah Karyawan Swasta.
2. Cluster 2 (C1)
   Cluster 2 dengan jumlah 36 data siswa didominasi Sekolah Islam (SDI) dan Sekolah Negeri (SDN), Untuk alamat didominasi Kedaung dengan jarak di bawah 5 KM dan pekerjaan rata-rata adalah Karyawan Swasta.
3. Cluster 3 (C2)

Creative Commons Attribution 4.0 International License

Cluster 3 dengan jumlah 76 data siswa didominasi Sekolah Islam (SDI), Untuk alamat didominasi Kedaung, Pamulang Barat dan Bambu Apus dengan jarak di bawah 5 KM dan pekerjaan rata-rata adalah Karyawan Swasta.

4. Cluster 4 (C3)
   Cluster 4 dengan jumlah 83 data siswa didominasi Sekolah Islam (SDI), Untuk alamat didominasi Kedaung, Ciputat dan Cipayung dengan jarak di atas 5 KM dan pekerjaan rata-rata adalah Karyawan Swasta.
5. Cluster 5 (C4)
   Cluster 5 dengan jumlah 56 data siswa didominasi Sekolah Islam (SDI), Untuk alamat didominasi Cipayung dan Ciputat dengan jarak di bawah 5 KM dan pekerjaan rata-rata adalah Karyawan Swasta, Wiraswasta.

2.1. Perbandingan Validasi Cluster

Hasil clustering yang didapat dari jumlah cluster K=3 dan K=5 dievaluasi menggunakan SSE (*Sum of Square Error*). Hasil pengklusteran terbaik apabila mempunyai nilai SSE (*Sum of Square Error*) terkecil.

Tabel 4.1 Hasil SSE (*Sum of Square Error*).

Jumlah Cluster	SSE (Sum of Square Error).
K=3	38,21077166311945
K=5	28,153419247762702

Berdasarkan tabel di atas menunjukkan nilai SSE (*Sum of Square Error*) yang paling rendah (hasil clustering terbaik) pada data siswa SMP Islam Al Syukro Universal adalah hasil clustering dari K=5 yang menghasilkan nilai SSE (Sum of Square Error) yaitu sebesar 28,153419247762702 dan melakukan iterasi sebanyak 9 kali.

KESIMPULAN

Berdasarkan hasil analisis dan pembahasan yang telah dilakukan, peneliti menyimpulkan beberapa hal penting terkait pengelompokan data siswa SMP Islam Al Syukro Universal menggunakan algoritma K-Means. Pada jumlah cluster K=3, diperoleh hasil pengelompokan yang terdiri dari cluster 1 (C0) dengan 95 data siswa, cluster 2 (C1) dengan 62 data siswa, dan cluster 3 (C2) dengan 140 data siswa. Sedangkan pada jumlah cluster K=5, data siswa terbagi ke dalam lima cluster yaitu cluster 1 (C0) sebanyak 46 siswa, cluster 2 (C1) sebanyak 36 siswa, cluster 3 (C2) sebanyak 76 siswa, cluster 4 (C3) sebanyak 83 siswa, dan cluster 5 (C4) sebanyak 56 siswa. Dari perbandingan kedua jumlah cluster tersebut, ditemukan bahwa pengelompokan dengan K=5 memberikan hasil evaluasi yang lebih baik dengan nilai SSE (Sum of Square Error) sebesar 28,153419247762702, yang menunjukkan bahwa model ini lebih optimal dalam mengelompokkan data siswa.

Selanjutnya, berdasarkan hasil clustering, peneliti mendapatkan bahwa target promosi calon siswa SMP Islam Al Syukro Universal dapat ditentukan dengan mempertimbangkan asal sekolah dan alamat siswa berdasarkan cluster yang telah

Creative Commons Attribution 4.0 International License

terbentuk serta jarak antara alamat siswa tersebut. Promosi dapat difokuskan pada cluster 1 (C0) yang mencakup Sekolah Dasar Islam (SDI) di wilayah Cipayung, Pamulang Barat, dan Pamulang Timur dengan jarak kurang dari 5 kilometer. Cluster 2 (C1) mencakup Sekolah Dasar Islam (SDI) dan Sekolah Dasar Negeri (SDN) di wilayah Kedaung dengan jarak kurang dari 5 kilometer. Cluster 3 (C2) berisi Sekolah Dasar Islam di wilayah Kedaung, Pamulang Barat, dan Bambu Apus juga dengan jarak kurang dari 5 kilometer. Cluster 4 (C3) meliputi Sekolah Dasar Islam di wilayah Kedaung, Ciputat, dan Cipayung dengan jarak lebih dari 5 kilometer, sedangkan cluster 5 (C4) merupakan Sekolah Dasar Islam yang berada di wilayah Cipayung dan Ciputat dengan jarak kurang dari 5 kilometer. Dengan demikian, promosi yang efektif dapat difokuskan pada wilayah-wilayah tersebut sesuai dengan cluster untuk meningkatkan potensi penerimaan siswa baru di SMP Islam Al Syukro Universal.

Berdasarkan temuan tersebut, beberapa saran diberikan sebagai bahan pertimbangan untuk pengembangan lebih lanjut. Pertama, disarankan untuk meningkatkan kecepatan proses dalam sistem K-Means Clustering agar pengolahan data dapat berjalan lebih efisien dan responsif. Kedua, perluasan sumber data sangat dianjurkan karena penelitian ini baru mencakup data siswa SMP saja, sementara data dari jenjang TK, SD, SMP, dan SMA di lingkungan Islam Al Syukro Universal juga penting untuk dianalisis guna mendapatkan gambaran yang lebih komprehensif dalam penentuan strategi promosi. Ketiga, disarankan untuk melakukan penelitian lanjutan dengan menggunakan metode atau algoritma lain dalam proses pengelompokan agar dapat memperoleh hasil yang lebih optimal dan relevan dengan kebutuhan promosi sekolah di Perguruan Islam Al Syukro Universal. Dengan demikian, langkah-langkah tersebut diharapkan dapat membantu sekolah dalam mengembangkan strategi promosi yang lebih tepat sasaran dan efektif di masa mendatang.

DAFTAR PUSTAKA

Agustina, H., & Herdiansyah, M. I. (2020). The application of data mining by using K-Means clustering method in determining new students' admission promotion strategy. *International Journal of Engineering and Advanced Technology (IJEAT)*, 9(3). ISSN 2249–8958.

Asril, E., Wiza, F., & Yunefri, Y. (2015). Analisis data lulusan dengan data mining untuk mendukung strategi promosi Universitas Lancang Kuning. *Jurnal Teknologi Informasi & Komunikasi Digital Zone*, 6(2), 24–32.

Asroni, H. F., & Prasetyo, E. (2018). Penerapan metode clustering dengan algoritma K-Means pada pengelompokkan data calon mahasiswa baru di Universitas Muhammadiyah Yogyakarta (Studi kasus: Fakultas Kedokteran dan Ilmu Kesehatan, dan Fakultas Ilmu Sosial dan Ilmu Politik). *SEMESTA TEKNIKA*, 21(1), 60–64. https://doi.org/10.18196/st.211211

Astria, C., Windarto, A. P., Wanto, A., & Irawan, E. (2019). Metode K-Means pada pengelompokan wilayah pendistribusian listrik. *Seminar Nasional Sains dan Teknologi Informasi (SENSASI)*. ISBN 978-602-52720-2-8.

Astuti, D., Iskandar, A. R., & Febrianti, A. (2019). Penentuan strategi promosi usaha mikro kecil dan menengah (UMKM) menggunakan metode CRISP-DM dengan

Creative Commons Attribution 4.0 International License

algoritma K-Means clustering. *Journal of Informatics, Information System, Software Engineering and Applications*, 1(2), 60–72. ISSN 2622-8106.

Basiroh. (2020). Sistem pendukung keputusan penentuan kelas promosi menggunakan K-Means clustering. *Prosiding Seminar Nasional NCIET*, 1, C25–C34. e-ISSN 2746-0975.

Budiman, R., & Rudianto. (2019). Penerapan data mining untuk menentukan lokasi promosi penerimaan mahasiswa baru pada Universitas Banten Jaya (Metode K-Means clustering). *Jurnal ProTekInfo*, 6(1). ISSN 2406-7741.

Fajriah, R. I., Sutisna, H., & Simpony, B. K. (2019). Perbandingan distance space Manhattan dengan Euclidean pada K-Means clustering dalam menentukan promosi. *IJCIT (Indonesian Journal on Computer and Information Technology)*, 4(1), 36–49. p-ISSN 2527-449X, e-ISSN 2549-7421.

Hartati, T., Nurdiawan, O., & Wiyandi, E. (2021). Analisis dan penerapan algoritma K-Means dalam strategi promosi kampus Akademi Maritim Suaka Bahari. *Jurnal Sains Teknologi Transportasi Maritim*, 3(1). e-ISSN 2722-1679, p-ISSN 2684-9135.

Hutabarat, J. T. (2018). Penerapan algoritma K-Means untuk menentukan lokasi promosi produk minuman penurun kolesterol Nutrive Benecol pada PT. Perkasa/Kalbe Nutritionals. *Jurnal Pelita Informatika*, 7(2). ISSN 2301-9425.

Julianto, V., Permadi, J., & Noviyanti. (2017). Aplikasi pemilihan strategi promosi penerimaan mahasiswa baru Politeknik Negeri Tanah Laut menggunakan metode K-Means clustering. *Jurnal Ilmiah Informatika*, 2(1).

Kasri, M. A., & Jati, H. (2020). Combination of K-Means and simple additive weighting in deciding locations and strategies of university marketing. Submitted June 25, 2020; Revised August 10, 2020; Accepted September 30, 2020.

Kesuma, F. T., & Feryanto. (2019). Penerapan data mining untuk menentukan penjualan sparepart Toyota dengan metode K-Means clustering. *Jurnal Sistem Informasi Ilmu Komputer Prima (JUSIKOM PRIMA)*, 2(2). e-ISSN 2580-2879.

Kurniawan, A., & Zein, A. (2024). Implementasi Data Mining Dalam Memprediksi Persediaan Cabinet Furniture Menggunakan Metode Algoritma Apriori Pada PT. Karya Nadiso Utama. *Jurnal Informatika Utama*, *2*(2), 81-96.

Kusuma, D. T., & Agani, N. (2015). Prototipe komparasi model clustering menggunakan metode K-Means dan FCM untuk menentukan strategi promosi: Studi kasus Sekolah Tinggi Teknik-PLN Jakarta. *Jurnal TICOM*, 3(3). ISSN 2302-3252.

Marajari, M. R., Nasution, K., & Haramaini, T. (2021). Analisis data mahasiswa untuk mendukung strategi promosi pada Universitas Medan Area menggunakan algoritma K-Means clustering. ISBN 978-623-7297-39-0.

Muni, A. (2020). Analisis algoritma K-Means clustering untuk menentukan strategi promosi penjualan sepeda motor studi kasus PT. Alfa Scorpii. *JUTI-UNISI (Jurnal Teknik Industri UNISI)*, 4(1). ISSN 2614-1299, eISSN 2614-1302.

Prastyawan, Z. G., Bagaskara, M. R., & Fitriati, D. (2018). Segmentasi pelanggan restoran menggunakan metode clustering simple K-Means. *Seminar Nasional Informatika, Sistem Informasi dan Keamanan Siber (SEINASI-KESI)*, Jakarta, Indonesia, 1 Desember 2018.

Creative Commons Attribution 4.0 International License

Purwanti, T., Ramdhan, W., & Santoso, S. (2021). Penerapan metode klasterisasi K-Means untuk strategi promosi pada SMK Tamansiswa Sukadamai. *JUTSI (Jurnal Teknologi dan Sistem Informasi)*, 1(2), 141–146. ISSN 2774-9029 (Online).

Putra, R. R., & Wadisman, C. (2018). Implementasi data mining pemilihan pelanggan potensial menggunakan algoritma K-Means. *Journal of Information Technology and Computer Science*, 1(1). e-ISSN 2614-1574.

Ramadhani, A., Parini, P., & Fauziah, R. (2021). Clustering tingkat promosi kampus dengan menggunakan algoritma K-Means (studi kasus kampus STMIK Royal Kisaran). *Journal of Science and Social Research*, 4(1), 66–73. ISSN 2615-4307 (Print), ISSN 2615-3262 (Online).

Rismayadi, A. A., Fatonah, N. N., & Junianto, E. (2021). Algoritma K-Means clustering untuk menentukan strategi pemasaran di CV. Integreet Konstruksi. *Jurnal Responsif*, 3(1), 30–36. E-ISSN 2685-6964.

Rusli, M., Arifin, S. P., & Trisnadoli, A. (2017). Pengembangan sistem pendukung keputusan untuk penentuan lokasi promosi penerimaan mahasiswa baru. *Jurnal Komputer Terapan*, 3(1), 11–18.

Setiawan, R. (2016). Penerapan data mining menggunakan algoritma K-Means clustering untuk menentukan strategi promosi mahasiswa baru (studi kasus: Politeknik LP3I Jakarta). *Jurnal Lentera ICT*, 3(1). ISSN 2338-3143.

Sibarani, R., & Omby, O. (2018). Algoritma K-Means clustering strategi pemasaran penerimaan mahasiswa baru Universitas Satya Negara Indonesia. *Vol. I*(2), 44–50. p-ISSN 2620-620X, e-ISSN 2621-9840.

Silalahi, M. (2018). Analisis clustering menggunakan algoritma K-Means terhadap penjualan produk pada PT Batamas Niaga Jaya. *CBIS Journal*, 6(2), September 2018.

Silalahi, N. (2020). Penentuan strategi promosi Universitas Budi Darma menggunakan algoritma K-Means clustering. *Terapan Informatika Nusantara*, 1(1), 40–46. ISSN 2722-7987.

Supriyatna, S., & Fahrudin, E. (2024). Pemanfaatan Algoritma Text Mining dalam Menemukan Pola Risiko Bencana sebagai Pengetahuan Kebencanaan dari Dokumen Kajian Risiko Bencana (KRB). *Jurnal Informatika Utama*, *2*(1), 35-42.

Wahyuni, D., & Widodo. (2017). Implementasi algoritma K-Means clustering untuk mengetahui bidang skripsi mahasiswa multimedia pendidikan teknik informatika dan komputer Universitas Negeri Jakarta. *Jurnal PINTER*, 1(2). e-ISSN 2597-4475.

Yunita, F. (2018). Penerapan data mining menggunakan algoritma K-Means clustering pada penerimaan mahasiswa baru (studi kasus: Universitas Islam Indragiri). *Jurnal SISTEMASI*, 7(3), 238–249. E-ISSN 2540-9719.

Zufria, I. (2018). Penentuan potensi lokasi promosi calon mahasiswa baru perguruan tinggi swasta berbasis AHP (Analytical Hierarchy Process). *JISTech*, 3(1), Januari-Juni 2018. ISSN 2528-5718.

Zulfa, N. S. L., & Hadiana, A. (2021). Kajian data mining menggunakan algoritma K-Means dan K-Medoids dalam strategi promosi (studi kasus: Universitas Islam Al-Ihya Kuningan). *Jurnal Fakultas Teknik*, 2(2). e-ISSN 2746-220X, p-ISSN 2747-1209.

Creative Commons Attribution 4.0 International License