

Strategi Marketing Penerimaan Mahasiswa Baru Menggunakan *Machine Learning* dengan Teknik *Clustering*

Raditya Danar Dana^{1*)}, Cep Lukman Rohmat², Ade Rizki Rinaldi³

^{1,2,3}Program Studi Manajemen Informatika, STMIK IKMI, Cirebon

^{1,2,3}Jln. Perjuangan No. 10B Majasem Kota Cirebon, 45131, Indonesia

email: ¹radith_danar@yahoo.com, ²k_ceplukmanrahmat@gmail.com, ³belajarjava2gether@gmail.com

Special Issue on Seminar Nasional - Inovasi Dalam Teknologi Informasi & Teknologi Pembelajaran 2019

Abstract – The marketing activity of new student admissions is one of the efforts undertaken by a university to maintain its existence in order to remain known and gain interest from the wider community. From the results of observations made at the research location, marketing activities carried out so far are still carried out in the same way from year to year without distinguishing the characteristics of the target prospective registrants, so the marketing pattern undertaken is not necessarily effective for all prospective applicants who have different characteristics - different. Therefore, it is necessary to make an effort to target target applicants based on certain characteristics to facilitate the determination of strategies and marketing patterns for new student admissions. The aim of this research is to group students' spread data using Machine Learning Technology approach using Clustering technique. This research resulted in the grouping of registrants in the admission activities of new students divided into 3 cluster groups, namely cluster 1 by 11%, cluster 2 by 56% and cluster 3 by 33%.

Keywords– Machine Learning, Clustering, K-Means.

Abstrak – Kegiatan marketing penerimaan mahasiswa baru merupakan salah satu upaya yang dilakukan oleh sebuah Perguruan Tinggi untuk menjaga eksistensinya agar tetap dikenal dan mendapatkan minat dari masyarakat luas. Dari hasil observasi yang dilakukan dilokasi penelitian, kegiatan marketing yang dijalankan selama ini masih dilakukan dengan cara yang sama dari tahun ketahun tanpa membedakan karakteristik dari target calon pendaftar, sehingga pola marketing yang dilakukan belum tentu efektif bagi seluruh calon pendaftar yang memiliki karakteristik yang berbeda – beda. Oleh karena itu perlu dilakukan sebuah upaya pengelompokan target calon pendaftar berdasarkan kareakteristik tertentu untuk memudahkan penentuan strategi dan pola marketing penerimaan mahasiswa baru. Penelitian ini bertujuan untuk melakukan pengelompokan terhadap data pesebaran mahasiswa dengan pendekatan Teknologi Machine Learning dengan menggunakan teknik Clustering. Penelitian ini menghasilkan pengelompokkan pendaftar dalam kegiatan penerimaan mahasiswa baru yang dibagi menjadi 3 kelompok cluster, yaitu cluster 1 sebanyak 11%, cluster 2 sebanyak 56 % dan cluster 3 sebanyak 33 %.

Kata Kunci – Machine Learning, Clustering, K-Means.

*) penulis korespondensi: Raditya Danar Daa
Email: radith_danar@yahoo.com

I. PENDAHULUAN

Fakultas Kedokteran Universitas Swadaya Gunung Jati, merupakan salah satu lembaga pendidikan perguruan tinggi di Cirebon yang setiap tahunnya secara rutin melakukan kegiatan marketing untuk penerimaan calon mahasiswa baru. Dalam menjalankan kegiatan marketing disetiap periode penerimaan mahasiswa baru, pihak pengelola masih sering menghadapi permasalahan yaitu belum diketahuinya pola atau strategi marketing yang tepat. Pola – pola marketing yang dilakukan selama ini belum menunjukkan pertumbuhan yang signifikan terhadap jumlah pendaftar bahkan berbanding terbalik dengan biaya marketing yang dikeluarkan setiap tahunnya yang cenderung meningkat

Guna menentukan strategi marketing yang tepat, diperlukan serangkaian analisa data untuk mengetahui karakteristik dari target customer. Teknik clustering merupakan bagian dari disiplin ilmu *Machine Learning* yang sangat efisien dalam melakukan analisa kelompok – kelompok data untuk mengetahui karakteristik masing – masing kelompok data customer [1]. Dengan diketahuinya karakteristik customer maka strategi marketing akan lebih mudah untuk dirumuskan sehingga tepat sasaran[2].

Dalam penelitian ini akan dilakukan analisis *cluster* berdasarkan data karakteristik pendaftar yang diperoleh dari kegiatan penerimaan mahasiswa baru di tahun – tahun sebelumnya. Analisis cluster akan dilakukan berdasarkan dua parameter utama yaitu “Kota pendaftar” dan “Asal Informasi”, berdasarkan dua informasi tersebut akan dilakukan proses clustering untuk mendapatkan informasi baru (knowledge discovery), berupa hubungan antara “Kota Pendaftar” dengan “Asal Informasi” sehingga akan lebih mudah untuk menentukan pola marketing yang tepat untuk target calon pendaftar dengan karakteristik tertentu.

Analisis cluster akan dilakukan dengan menggunakan teknologi *Machine Learning* dengan teknik Clustering metode *K-Mean*. Metode *K-Mean* digunakan secara luas dalam teknik clustering, *K-Mean* memiliki keunggulan dalam kecepatan komputasi dan tahapan-tahapan proses yang relatif mudah [3].

II. PENELITIAN TERKAIT

Somantri, O., dkk, melakukan penelitian *metode k-means* untuk optimasi klasifikasi tema tugas akhir mahasiswa menggunakan *Support Vector Machine (SVM)*. Penelitian

yang dilakukannya ini membahas tentang menentukan *klasifikasi* tema dari tugas akhir mahasiswa yang sering kali dialami di setiap perguruan tinggi. *Algoritma SVM dan K-Means* digunakan untuk mengklasifikasikan berbagai jenis tema Tugas Akhir mahasiswa. Dataset yang digunakan adalah data jurnal mahasiswa D3 Teknik Komputer Politeknik Harapan Bersama dengan jumlah sampel sebanyak 131 jurnal mahasiswa. Hasilnya algoritma SVM dan K-Means dapat mengoptimalkan dan meningkatkan akurasi yang lebih baik yaitu 86,21% [4].

Penelitian yang dilakukan Ismiatun, Siti & Wakhidah, Nur terkait identifikasi pengelompokkan kondisi permukaan jalan menggunakan *algoritma K-Means*. Penelitian ini membahas tentang pengelompokkan kondisi permukaan jalan bertujuan untuk mengetahui kecamatan mana yang memiliki kondisi jalan paling rawan rusak karena jalan yang rusak sering kali menjadi pemicu kecelakaan lalu lintas. Pengelompokkan ini menggunakan *algoritma K-Means* dimana menghasilkan 7 kluster dengan jumlah data 58 sample dengan jumlah atribut 4. Terdapat satu kecamatan yang paling rawan jalan rusak yaitu kecamatan Gayamsari[5].

Berikutnya adalah penelitian dari Yusuf, F., dkk, melakukan penelitian optimasi *k-means clustering* menggunakan *particle Swarm Optimization* pada sistem identifikasi tumbuhan obat berbasis citra. Penelitian ini bertujuan untuk membangun sistem yang mengidentifikasi tumbuhan obat dengan teknik *klustering* dengan mengelompokkan data citra yang sesuai berdasarkan spesies tumbuhan obat. Metode *Particle Swarm Optimization* digunakan untuk mengatasi kelemahan metode *K-Means* yaitu menentukan pusat *cluster* awal. Analisis ini dilakukan untuk 30 jenis spesies tumbuhan obat yang ada di Indonesia dengan jumlah 48 dari tiap spesiesnya. Hasilnya *Particle Swarm Optimization* dapat meningkatkan kinerja dari metode *K-Means* dalam proses identifikasi tumbuhan obat dengan akurasi *K-Means* tanpa *Particle Swarm Optimization* diperoleh 31.67%, sedangkan untuk *K-means* dengan *Particle Swarm Optimization* diperoleh 47.33% [6].

Penelitian Sudrajat, Dadang., dkk, meneliti tentang sistem pembelajaran *blended learning* dengan 6 penilaian dapat meningkatkan kualitas pembelajaran. Hanya saja tingkat kepuasan berdasarkan 6 nilai memiliki nilai yang bervariasi. Tujuan penelitian ini bertujuan mengelompokkan kepuasan dalam sistem pembelajaran *blended learning* dengan 6 nilai adaptif kompleks. Objek penelitian ini bertempat di 3 SMK Kota Cirebon. Data kepuasan dikelompokkan dengan menggunakan metode *K-Means*. Nilai *K* yang digunakan sebesar 4, dan menghasilkan *Davies Bouldiens Index* sebesar -1.049 [7].

III. METODE PENELITIAN

A. Data Collection

Data diperoleh dari kegiatan penerimaan mahasiswa baru pada tahun 2017 yang berjumlah 184 pendaftar. Data tersebut akan dilakukan proses *pre-processing* terlebih dahulu untuk memastikan semua data telah terisi atau tidak ada data yang *Null*. *Pre-processing* juga dimaksudkan untuk mengkonversi data yang masih berupa label dijadikan numerik karena pada komputasi *clustering* hanya dapat melakukan pemrosesan data yang bersifat numerik. Berikut dibawah ini data label yang menjadi fokus penelitian:

- Nama Pendaftar, yaitu nama lengkap pendaftar atau calon mahasiswa baru.
- Kota tempat asal pendaftar, yaitu kota asal pendaftar yang tersebar berasal dari 40 wilayah yang berbeda beda.
- Sumber informasi penerimaan mahasiswa baru, yaitu darimana pendaftar mendapatkan sumber informasi penerimaan mahasiswa baru yang dikelompokkan kedalam 4 katagori, “Kolega”, “Iklan Elektronik”, “Promosi Ke Sekolah” dan “media iklan non elektronik”.

Proses konversi data label menjadi numerik dilakukan dengan mengurutkan data berdasarkan banyaknya frekuensi kemunculan. Pada Tabel I adalah hasil konversi label “Kota tempat asal pendaftar” menjadi nilai numerik diurutkan 10 data terbanyak berdasarkan frekuensi tingkat kemunculan pendaftar dari kota tertentu. Pada Tabel II adalah hasil konversi label “Sumber Informasi Penerimaan Mahasiswa Baru”, menjadi nilai numerik diurutkan berdasarkan frekuensi tingkat kemunculan asal sumber informasi pendaftar

TABEL I
 KONVERSI LABEL “ASAL KOTA” MENJADI NUMERIK

No	Kota	Frekuensi	Inisial
1	Cirebon	39	40
2	Kuningan	14	39
3	Subang	10	38
4	Indramayu	9	37
5	Tangerang	9	36
6	Bogor	8	35
7	Brebes	8	34
8	Bandung	6	33
9	Karawang	6	32
10	Majalengka	6	31

TABEL II
 KONVERSI LABEL “SUMBER INFORMASI” MENJADI NUMERIK

No	Kota	Frekuensi	Inisial
1	Kolega	115	4
2	Iklan Elektronik	62	3
3	Promosi Sekolah	4	2
4	Iklan Non Elektronik	3	1

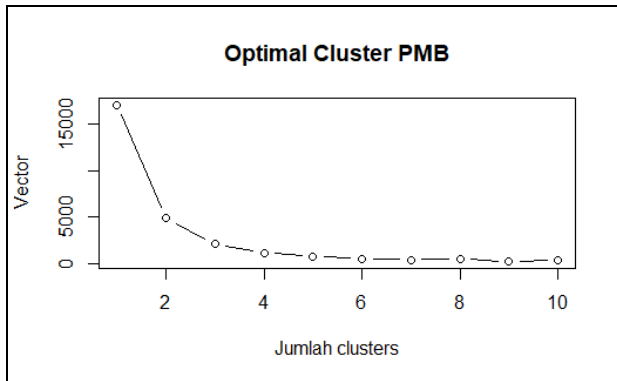
B. Menentukan Jumlah Cluster

Tahapan ini pada prinsipnya adalah sebuah upaya untuk menentukan jumlah cluster yang optimal. Penentuan jumlah cluster merupakan bagian yang sangat penting dalam proses pengelompokkan data pada *machine learning* dan akan sangat berimplikasi terhadap hasil penelitian[8], jumlah cluster yang terlalu sedikit tidak mencerminkan keragaman data sedangkan jumlah cluster yang terlalu banyak akan membebani proses komputasi dan menyulitkan dalam mendeskripsikan karakteristik data masing-masing *cluster* karena perbedaan karakteristik data antara sebuah *cluster* dengan *cluster* yang lain menjadi tidak terlalu signifikan [9].

Banyak jenis metode yang dapat digunakan untuk menentukan jumlah *cluster*, pada penelitian ini akan digunakan metode *Elbow* karena metode *elbow* menawarkan

komputasi yang cepat dan efisien untuk jenis data yang tidak terlalu kompleks [10].

```
ZCSS <- vector()
for (i in 1:10) ZCSS[i] <-
sum(kmeans(X_pmb,i)$withinss)
plot(1:10, ZCSS, type = "b", main =
paste('Clusters of pmb'), xlab = 'Number od
clusters', ylab = "ZCSS")
```



Gbr. 1 Visualisasi Optimal Cluster (Software R.Studio)

Melalui proses visualisasi yang dihasilkan seperti pada Gbr.1 menggunakan metode *Elbow* diperoleh gambaran bahwa jumlah cluster yang optimal adalah 3, karena mulai dari 4 dan seterusnya tidak memberikan perubahan yang cukup signifikan.

C. *Experimen Dan Pengujian Algoritma*

Algoritma yang digunakan untuk pembentukan cluster adalah algoritma *K-Means*. *K-Means* merupakan salah satu metode data non-hierarchical clustering yang dapat mengelompokkan data mahasiswa kedalam beberapa cluster berdasarkan kemiripan dari data tersebut, sehingga data mahasiswa yang memiliki karakteristik yang sama dikelompokkan dalam satu cluster dan yang memiliki karakteristik yang berbeda dikelompokkan dalam cluster yang lain. Implementasi menggunakan R Programming Language digunakan untuk membantu menemukan nilai yang akurat. Atribut yang digunakan adalah kota asal mahasiswa. Hasil dari penelitian ini digunakan sebagai salah satu dasar pengambilan keputusan untuk menentukan strategi promosi berdasarkan cluster yang terbentuk oleh panitia promosi fakultas kedokteran unswagati

Berikut ini adalah langkah – langkah dalam melakukan Clustering menggunakan algoritma *K-Means* [11].

- 1) Menentukan jumlah cluster berdasarkan kemungkinan nilai persepsi pasien.
- 2) Memilih objek yang akan dijadikan sebagai titik centroid masing – masing cluster.
- 3) Mengelompokkan objek ke titik centroid terdekat berdasarkan Euclidean Distance.
- 4) Menghitung ulang kembali semua nilai titik centroid.
- 5) Mengulangi kembali langkah 2 hingga nilai titik centroid tidak berubah.

Komputasi akan dilakukan menggunakan listing program menggunakan R.Studio sebagai berikut. Proses perhitungan metode *K-Means* dilakukan dengan cara mencari jarak minimum antara objek dengan masing – masing titik centroid setiap cluster [12].

```
kmeans_pmb <- kmeans(X_pmb,3, iter.max = 300,
nstart = 10)
```

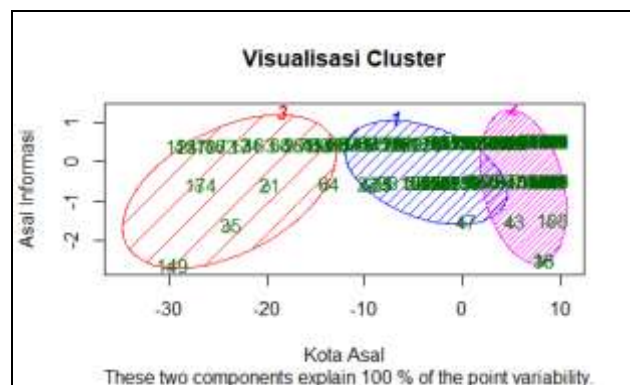
IV. HASIL DAN PEMBAHASAN

Proses perhitungan metode *K-Means* menghasilkan cluster seperti ditunjukkan pada Tabel III.

TABEL III
 CLUSTER YANG TERBENTUK

No	Cluster	Jumlah Item
1	Cluster 1	22
2	Cluster 2	103
3	Cluster 3	59
	Total Item	184

- 11% dari jumlah pendaftar penerimaan mahasiswa baru dikelompokkan kedalam “Cluster 1” dengan jarak rata rata titik object dengan cluster 1 adalah 3.5663.
- 56% dari jumlah pendaftar penerimaan mahasiswa baru dikelompokkan kedalam “Cluster 2” dengan jarak rata rata titik object dengan cluster 2 adalah 2.6724
- 33% dari jumlah pendaftar penerimaan mahasiswa baru dikelompokkan kedalam “Cluster 3” dengan jarak rata rata titik object dengan cluster 3 adalah 2.9312



Gbr. 2 Visualisasi pengelompokkan data (Software R.Studio)

Berikut dibawah ini adalah interpretasi karakteristik dari cluster yang terbentuk

- Cluster 1, didominasi oleh pendaftar yang berasal dari luar kota Cirebon dan luar pulau Jawa dengan sumber informasi yang diterima dominan sebagian besar berasal dari media elektronik.
- Cluster 2, didominasi oleh pendaftar berasal dari wilayah Cirebon dan sekitarnya dengan sumber informasi yang diterima dominan sebagian besar berasal dari kolega atau mulut kemulut dan media elektronik dan sebagian kecil berasal dari sosialisasi yang dilakukan ke sekolah sekolah.
- Cluster 3, didominasi oleh pendaftar berasal dari luar kota Cirebon namun masih di dalam wilayah Jawa

Barat dengan sumber informasi yang diterima bersumber dari kolega dan media elektronik.

Berdasarkan karakteristik *cluster* yang terbentuk maka dapat dirumuskan strategi marketing sebagai berikut:

- 1) Mengoptimalkan kegiatan marketing dengan melakukan promosi melalui media elektronik guna menjangkau calon – calon pendaftar yang berasal dari luar kota Cirebon
- 2) Mengoptimalkan kegiatan marketing dengan melakukan sosialisasi ke sekolah – sekolah guna menjangkau calon – calon pendaftar yang berasal dari dalam kota Cirebon
- 3) Kegiatan marketing yang dilakukan dengan menggunakan media non elektronik seperti brosur , baliho dan spanduk terbukti kurang efektif, sehingga perlu dikaji ulang pemanfaatannya agar tidak terlalu menyita pembiayaan marketing untuk kedepannya
- 4) Meningkatkan layanan pada stakeholder karena sumber informasi yang berasal dari kolega atau mulut kemulut terbukti sangat efektif dalam mempengaruhi calon pendaftar (Cluster 2) dengan karakteristik calon pendaftar yang berasal dari kota Cirebon dan sekitarnya

V. KESIMPULAN

Dari hasil eksperimen yang telah dilakukan dapat diketahui metode *Elbow* dapat digunakan untuk menentukan jumlah *cluster yang optimal dalam pengelompokan data penerimaan mahasiswa baru*, yaitu berjumlah 3 *cluster* yaitu “Cluster 1”, “Cluster 2” dan “Cluster 3”. Berdasarkan hasil analisis cluster menggunakan teknologi *Machine Learning* menggunakan metoda K-Means data penerimaan mahasiswa baru dapat dikelompokkan berdasarkan pola dan karakteristik tertentu dapat memudahkan penentuan strategi marketing yang optimal, efisien dari segi pembiayaan dan tepat sasaran.

UCAPAN TERIMA KASIH

Terimakasih kepada Yayasan STT Multimedia yang telah men-support dan turut berkontribusi pada penelitian ini.

DAFTAR PUSTAKA

- [1] C. Calvo-porrall and J. Lévy-mangin, “From ‘foodies ’ to ‘cherry-pickers ’: A clustered-based segmentation of specialty food retail customers,” *J. Retail. Consum. Serv.*, vol. 43, no. January, pp. 278–284, 2018.
- [2] C. P. Ezenkwo and S. Ozuomba, “Application of K-Means Algorithm for Efficient Customer Segmentation: A Strategy for Targeted Customer Services,” no. October, 2015.
- [3] S. Ghosh and S. K. Dubey, “Comparative Analysis of K-Means and Fuzzy C- Means Algorithms,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 4, no. 4, pp. 35–39, 2013.
- [4] O. Somantri and S. Wiyono, “Metode K-Means untuk Optimasi Klasifikasi Tema Tugas Akhir Mahasiswa Menggunakan Support Vector Machine (SVM),” vol. 3, no. 1, pp. 34–45, 2016.
- [5] S. Asmiatun and N. Wakhidah, “IDENTIFIKASI PENGELOMPOKKAN KONDISI PERMUKAAN JALAN MENGGUNAKAN ALGORITMA K-MEANS,” vol. 14, no. 1, pp. 17–23, 2018.
- [6] F. Y. Bisilisin, Y. Herdiyeni, and B. I. B. P. Silalahi, “Optimasi K-Means Clustering Menggunakan Particle Swarm Optimization pada Sistem Identifikasi Tumbuhan Obat Berbasis Citra K-Means Clustering Optimization Using Particle Swarm Optimization on Image Based Medicinal Plant Identification System,” vol. 3, no. 2002.
- [7] D. Sudrajat, R. D. Dana, N. Rahaningsih, A. R. Dikananda, and D. A. Kurnia, “Clustering Student s Satisfaction in Complex Adaptive Blended Learning with the Six Value System Using the K-Means Algorithm,” vol. 7, no. 9, pp. 1990–1995, 2019.
- [8] M. Nishom and M. Y. Fathoni, “Implementasi Pendekatan Rule-Of-Thumb untuk Optimasi Algoritma K-Means Clustering,” vol. 3, no. 2, pp. 237–241, 2018.
- [9] M. Azhar, J. Z. Huang, A. Masud, M. J. Li, and L. Cui, “A hierarchical Gamma Mixture Model-based method for estimating the number of clusters in complex data,” *Appl. Soft Comput. J.*, p. 105891, 2019.
- [10] Dibya Jyoti Bora and D. A. K. Gupta, “A Comparative study Between Fuzzy Clustering Algorithm and Hard Clustering Algorithm,” *Int. J. Comput. Trends Technol.* –, vol. 10, no. 2, pp. 108–113, 2014.
- [11] D. T. Kusuma, N. Agani, J. Ticom, and V. No, “Prototipe Komparasi Model Clustering Menggunakan Metode K-Means Dan FCM untuk Menentukan Strategi Promosi : Study Kasus Sekolah Tinggi Teknik-PLN Jakarta,” vol. 3, no. 3, pp. 1–10, 2015.
- [12] T. Bucker, “Customer Clustering in the Insurance Sector by Means of Unsupervised Machine Learning,” 2016.