

ARTIKEL

**PENERAPAN METODE *CLUSTERING* UNTUK REKOMENDASI
PEMILIHAN MOBIL BEKAS DENGAN ALGORITA *K-MEANS***



Oleh:

LUTFI ARI SUNGGARA

14.1.03.03.0110

Dibimbing oleh :

- 1. Sucipto, M.Kom.**
- 2. Teguh Andriyanto, ST., M.Cs.**

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNIK
UNIVERSITAS NUSANTARA PGRI KEDIRI**

2019



SURAT PERNYATAAN ARTIKEL SKRIPSI TAHUN 2019




Yang bertanda tangan dibawah ini:

Nama Lengkap : Lutfi Ari Sunggara
NPM : 14.1.03.03.0110
Telepon/HP : 082257165057
Alamat Surel (Email) : lutfi02011994@gmail.com
Judul Artikel : PENERAPAN METODE *CLUSTERING* UNTUK
REKOMENDASI PEMILIHAN MOBIL BEKAS
DENGAN ALGORITMA *K-MEANS*
Fakultas – Program Studi : Teknik – Sistem Informasi
Nama Perguruan Tinggi : Universitas Nusantara PGRI Kediri
Alamat Perguruan Tinggi : Jl. KH.Ahmad Dahlan No.76

Dengan ini menyatakan bahwa:

- artikel yang saya tulis merupakan karya saya pribadi (bersama tim penulis) dan bebas plagiarisme;
- artikel telah diteliti dan disetujui untuk diterbitkan oleh Dosen Pembimbing I dan II.

Demikian surat pernyataan ini saya buat dengan sesungguhnya. Apabila dikemudian hari ditemukan ketidak sesuaian data dengan pernyataan ini dan atau ada tuntutan dari pihak lain, saya bersedia bertanggung jawab dan diproses sesuai dengan ketentuan yang berlaku.

Mengetahui		Kediri, 30 Januari 2019
Pembimbing I  Sucipto, M.Kom. NIDN. 0725057003	Pembimbing II  Teguh Andriyanto, ST., M.Cs. NIDN. 0721029101	Penulis,  Lutfi Ari Sunggara NPM. 14.1.03.03.0110

PENERAPAN METODE *CLUSTERING* UNTUK REKOMENDASI PEMILIHAN MOBIL BEKAS DENGAN ALGORITMA *K-MEANS*

Lutfi Ari Sunggara

14.1.03.03.0110

Teknik – Sistem Informasi

Lutfi02011994@gmail.com

Sucipto, M.Kom., Teguh Andriyanto, ST., M.Cs.

UNIVERSITAS NUSANTARA PGRI KEDIRI

ABSTRAK

Dengan banyaknya pilihan jenis dan *tipe* mobil yang ditawarkan tersebut tak jarang banyak konsumen yang merasa bingung dalam memilih jenis dan *tipe* mobil apa yang sesuai kebutuhannya, masalah seperti ini banyak dialami oleh para konsumen yang bingung dengan banyaknya pilihan yang ada. Penelitian ini menggunakan algoritma *k-means* untuk mengelompokkan mobil berdasarkan spesifikasi dengan bahasa pemrograman PHP. Algoritma *k-means* dapat digunakan untuk mengelompokkan data mobil untuk membantu memberikan rekomendasi pemilihan mobil. Dari hasil *clustering* yang didapatkan pada jumlah 3 *cluster* yaitu *cluster*1 kategori murah, *cluster* 2 kategori sedang, dan *cluster* 3 kategori mahal. Kategori murah ditujukan kepada pembeli yang memilih dengan harga yang rendah. Kategori sedang ditujukan kepada pembeli yang menyesuaikan mobil dengan kondisi yang lumayan baik dan harga yang tidak terlalu tinggi. Untuk kategori mahal disesuaikan dengan keinginan pembeli yang mempunyai selera tinggi untuk kelas-kelas mobil mewah.

Kata Kunci — *Data Mining, Clustering, K-Means*

1. LATAR BELAKANG

Kendaraan adalah alat transportasi yang digerakkan oleh mesin maupun oleh makhluk hidup. Kendaraan tersebut seperti mobil, motor, kereta, perahu, dan pesawat, tetapi ada juga kendaraan yang tidak bermesin yaitu kendaraan yang digerakkan oleh manusia atau ditarik oleh hewan, seperti gerobak. Semua kendaraan tersebut adalah alat transportasi yang sudah dianggap sebagian besar masyarakat sebagai kebutuhan yang penting. Peranan transportasi sebagai mobilitas atau perpindahan manusia atau barang dari suatu tempat ketempat yang lainnya.

2. METODE PENELITIAN

Metode penelitian ini menggunakan metode penelitian *Action Research* (Penelitian Tindakan) yaitu penelitian yang berfokus langsung pada suatu tindakan, dengan memahami dan mencatat pola-pola yang ada. Secara metodologis tidak kuat namun ada *knowledge* atau pengetahuan yang bias digali. Metode penelitian *Action Research* memiliki 4 tahapan penelitian yakni Perencanaan Tindakan (*Action Plan*), Pelaksanaan Tindakan (*Action Taking*) Pengamatan atau evaluasi, dan Refleksi (*Reflection*). Dari keempat tahapan tersebut membentuk suatu siklus. Dimana satu siklus jika penelitian ini belum sesuai dengan harapan atau

terjadinya masalah maka akan mengulangi tahapan-tahapan pada siklus berikutnya sampai mencapai tujuan yang diharapkan dari adanya pelaksanaan penelitian ini.

Data dalam penelitian ini diambil dari situs <http://kaggle.com>. Kaggle adalah salah satu komunitas terbesar ilmuwan data (data scientist) diseluruh dunia. Di Kaggle ini, para data scientist bisa berkompetisi untuk menyelesaikan permasalahan ilmiah berdasarkan data yang kompleks. Data yang diperoleh dalam penelitian ini merupakan data sekunder, yaitu data dari sumber yang telah ada. Data yang digunakan dalam penelitian ini Used Cars Database yang telah dikumpulkan oleh Orges Leka yang diambil dari situs <https://www.kaggle.com/orgesleka/used-cars-database>.

2.1 Analisis Data

Pemrosesan data mining dalam penelitian ini akan digunakan metode clustering. Metode clustering termasuk dalam unsupervised learning dimana tidak ada pendefinisian kelas ataupun label dalam dataset. Sehingga clustering dapat digunakan untuk menentukan label kelas bagi data-data yang belum diketahui kelasnya.

Algoritma pada metode clustering yang akan dipakai dalam penelitian ini adalah algoritma K-

Means. Algoritma K-Means merupakan metode clustering non hierarki yang memiliki waktu komputasi yang relatif cepat. Berdasarkan analisis perbandingan antara K-Means dan Fuzzy C-Means (FCM) yang dilakukan oleh Suomi G. dan Sanjay Kumar D., didapatkan hasil yang membuktikan bahwa algoritma K-Means lebih cepat dengan elapsed time 0.433755 detik dibandingkan dengan algoritma FCM yang memiliki elapsed time sebesar 0.781679 detik. (Ridlo, Defiyanti, & Primajaya, 2017)

Alat yang digunakan dalam penelitian ini ialah menggunakan tools Microsoft Excel. Microsoft Excel digunakan sebagai perhitungan manual dari setiap langkah-langkah pada proses mining. Dimana proses mining yang dimulai dari seleksi data, pemrosesan awal (Preprocessing), Pengklasteran data, hingga menghasilkan kluster data serta melakukan validasi dan pengujian.

2.2 Data Mining

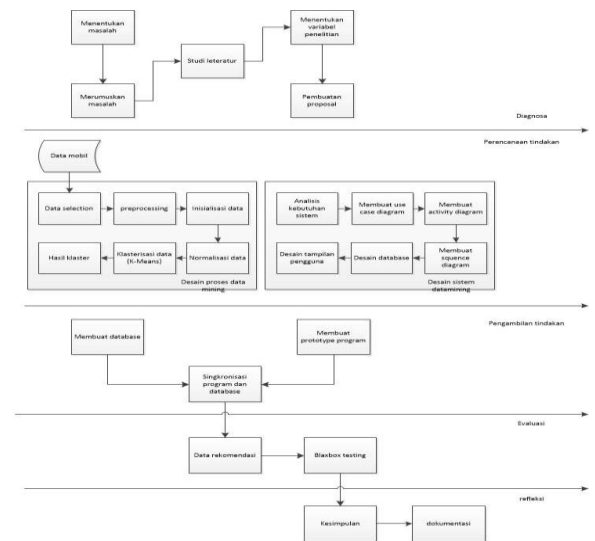
Pemrosesan *data mining* dalam penelitian ini akan digunakan metode *clustering*. Metode *clustering* termasuk dalam *unservised learning* dimana tidak ada pendefinisian kelas ataupun label dalam dataset. Sehingga *clustering* dapat digunakan untuk menentukan label kelas bagi data-data yang belum diketahui kelasnya.

Cluster adalah kumpulan catatan yang mirip satu sama lain, dan berbeda dengan catatan dalam kelompok lain. Clustering berbeda dari klasifikasi karena tidak ada variabel target untuk clustering. Tugas pengelompokan tidak mencoba mengklasifikasikan, memperkirakan, atau memprediksi nilai variabel target. Sebagai gantinya, algoritma pengelompokan berusaha untuk mengelompokkan keseluruhan data ke dalam subkelompok atau cluster yang relatif homogen, di mana kesamaan catatan dalam cluster dimaksimalkan dan kesamaan dengan catatan di luar cluster diminimalkan.

K-means clustering merupakan salah satu metode data clustering non-hirarki yang mengelompokkan data dalam bentuk satu atau lebih cluster atau kelompok. Data-data yang memiliki karakteristik yang sama dikelompokkan dalam satu cluster atau kelompok dan data yang memiliki karakteristik yang berbeda dikelompokkan dengan cluster atau kelompok yang lain sehingga data yang berada dalam satu cluster atau kelompok memiliki tingkat variasi yang kecil

2.3 Alur Penelitian

Dalam penelitian ini menggunakan alur penelitian sebagai berikut :



Gambar 2.1 Alur Penelitian

Pada tahapan diagnosa terdapat 5 proses yakni menentukan masalah, merumuskan masalah, studi literatur, menentukan variabel penelitian, dan pembuatan proposal. Dimana pada tahapan diagnosa ini adalah menentukan garis besar dari penelitian ini.

Tahap perencanaan ini dibagi menjadi 2 kelompok yakni desain proses mining dan desain sistem mining. Pada tahap perencanaan ini terdapatnya data kecelakaan yang akan dilakukan pemrosesan data mining. pada desain proses mining terdapat proses-proses perencanaan untuk melakukan klasterisasi dari data kecelakaan, dimana proses-prosesnya ialah : *data selection*, *preprocessing*, *data transformation*, klasterisasi data (*K-Means*), hasil *cluster*, serta validasi dan pengujian. Desain sistem mining terdapat proses yakni melakukan analisis kebutuhan sistem, membuat *use case diagram*, membuat *activity diagram*, membuat *sequence diagram*, desain database, dan desain tampilan pengguna.

Pada tahapan pelaksanaan ini merupakan proses pelaksanaan dari tahap perencanaan. Dalam tahap ini, yang dilakukan adalah membuat database dan membuat prototype program, yang selanjutnya akan dilakukan sinkronisasi antara database dengan program Pada prototype program tersebut terdapat proses mining yakni pengklasteran (*Clustering*) dengan algoritma *K-Means Clustering*.

Tahapan pengamatan merupakan tahapan dimana dilakukannya evaluasi terhadap pelaksanaan dari penelitian ini. Pada tahap ini kluster data akan dilakukan uji validasi dengan *Silhouette Coefficient* yang selanjutnya akan menghasilkan sebuah data rekomendasi. Untuk pengujian program akan diujikan dengan *Black-box Testing*.

Tahapan yang terakhir ialah tahap refleksi, dimana pada tahap ini program dapat

diimplementasikan dengan baik. Selanjutnya dapat diperoleh kesimpulan dari implementasi program tersebut, guna melaksanakan dokumentasi atas semua proses penelitian.

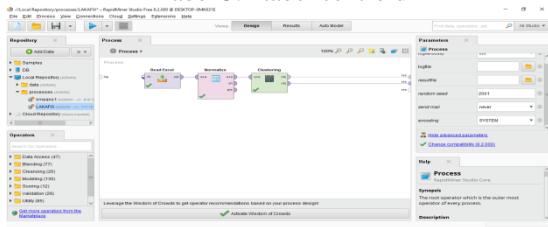
3. HASIL DAN PEMBAHASAN

3.1. Proses Mining

Tahap pertama dari algoritma *K-Means* yaitu menentukan pusat *cluster* awal. Menentukan *centroid* awal dilakukan secara acak dari data atau objek yang tersedia sebanyak jumlah *cluster* *k*. Nilai *centroid* awal pada penelitian ini dilakukan pemilihan secara acak dengan contoh perhitungan sebagai berikut :

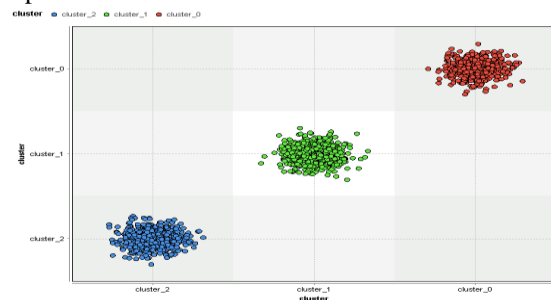
ID	cluster	price	vehicleType	yearOfRegis...	gearbox	powerPS	kilometer	fuelType
15	cluster_0	0.197	0.500	0.929	1	0.010	0.333	0.333
16	cluster_1	0.190	0.500	0.714	0	0.007	1	0.333
17	cluster_2	0.132	0.750	0.714	0	0.001	0	0.333

Tabel 3.1 tabel centroid



Gambar 3.2 Model Pengujian RapidMiner

Pada aplikasi RapidMiner pertama yang dilakukan adalah mengimport data yang akan diolah kedalam RapidMiner. Kemudian pada tabel operasi dipilih modeling k-means untuk metode clustering. Hubungkan tiap tool tersebut kemudian pilih tombol run dan otomatis data akan diolah kedalam metode clustering. Didapatkan hasil seperti dibawah.



Gambar 3.3 Hasil Sebaran Centroid

Terdapat sebaran 3 centroid yang sudah ditentukan peneliti. Dan juga dapat dilihat hasil dari perhitungan yaitu nilai cluster yang didapat seperti pada gambar dibawah.

Attribute	cluster_0	cluster_1	cluster_2
price	0.212	0.109	0.128
vehicleType	0.583	0.489	0.495
yearOfRegistration	0.630	0.542	0.624
gearbox	1	0	0
powerPS	0.018	0.007	0.009
kilometer	0.579	0.834	0.165
fuelType	0.162	0.109	0.098

Gambar 3.4 Nilai Centroid

3.2. Pengujian Data

Pada pengujian kali ini peneliti menggunakan langkah pengujian manual yang diterapkan pada Microsoft Excel dengan rumus manual untuk mencari nilai centroid pada kumpulan data tertentu.

Kemudian rumus tersebut diimplementasikan kedalam tabel excel untuk mencari nilai centroid pada tiap klaster yang sudah ditentukan yaitu :

$$=SQRT(((C3-U2)^2)+((D3-V2)^2)+((E3-W2)^2))$$

Setelah diketahui nilai pada tiap klaster tersebut baru dihitung untuk menentukan nilai centroid pusat untuk tiap data yang dicari yaitu:

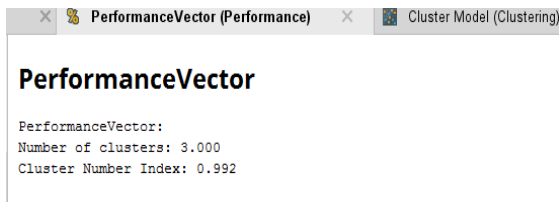
$$=IF(H3<K3;IF(H3<N3;"1";"3");IF(K3<H3;IF(K3<N3;"2";"3")))$$

Berikut ini merupakan hasil pengujian sistem yang telah dibuat oleh peneliti.

No	Nama mobil	Power PS	Kilometer	Price	Centroid 1			Centroid 2			Centroid Price	
					100	150000	650	100	150000	5000		170
1	Bmw_316i_compact_2004	100	150000	200		450,00		4800,00			81399,36	1
2	Bmw_316i	60	150000	200		451,96		4800,25			81399,43	1
3	Vw_golf_3_Size_Medekonstruktio	90	150000	300		390,21		4700,04			81312,36	1
4	Volkswagen_Bea_1180_101_70	122	150000	350		300,57		4650,00			81312,36	1
5	VolKswagen_Bea_1180_101_70	88	150000	350		300,33		4650,05			81312,42	1
6	VW_Polo_Coupe_86C_101_101_101	54	125000	370		290,61		25425,16			56887,03	1
7	Opel_Corva_3_101_101	45	150000	444		212,74		4556,45			81295,72	1
300	Audi_A8_A8_101_101_101	129	150000	11900		11520,00		6900,00			80056,24	2
301	Volkswagen_Polo_11_70_202_MAR74	90	30000	11919		11952,97		120169,30			40111,20	2
302	VW_Tiguan_31_70_201_201	101	150000	12000		11520,00		7000,00			80051,57	2
303	Bmw_520i_520i	121	150000	11900		11850,70		7500,00			80056,02	2
304	Bmw_530i_Touring	258	150000	11950		11001,01		7851,45			80051,69	2
305	Audi_A8_A8_101_101_101	140	150000	11999		11949,00		7889,00			80051,03	2
306	Audi_A8_A8_101_101_101	179	150000	11999		11949,04		7999,01			80051,59	2
307	Bmw_M1_11_70_Ambition	105	40000	11900		11971,61		11900,00			10046,10	2
308	Bmw_200i_Touring	170	200000	11940		5300,07		30700,70			30071,49	3
309	Volkswagen_Golf_Variant_14_70_201_Highline_Merou_Panorama	160	60000	11949		9001,00		30400,40			10007,87	3
310	Volkswagen_Bea_1180i_Automatik	109	150000	11900		11950,00		6500,00			80011,27	2
311	Bmw_116d_116d	140	80000	11949		71203,31		7057,11			10055,11	3
312	Audi_A4_A4_11_70_Merou_Schneid_51_101_TOP_Justand	100	60000	14900		90209,24		30479,22			10018,11	3
313	Volkswagen_Coupe_100_Coupe_Automatik_Aerospange	170	125000	14949		10570,70		10143,67			15001,40	3
314	Ford_C_Merou_Titanium_101_101_Estimate	125	30000	14900		100796,61		10079,46			40000,00	3
315	Audi_A4_11_70_101	81	60000	14900		91024,41		36500,00			10000,00	3

Gambar 3.5 Model Pengujian Manual Excel

Pengujian juga dilakukan pada aplikasi RapidMiner yang menggunakan tool-tool yang sudah disediakan pada menu aplikasi.



Gambar 3.6 Model Pengujian RapidMiner

Pada gambar 3.6 menunjukkan bahwa pada pengujian aplikasi RapidMiner terdapat 3 kluster yang sudah ditentukan dan memiliki nilai kluster 0.992 dan dapat dibulatkan menjadi 1.

4. SIMPULAN

Algoritma *k-means* dapat digunakan untuk mengelompokkan data mobil untuk membantu memberikan rekomendasi pemilihan mobil. Dari hasil *clustering* yang didapatkan pada penerapan *data mining* dengan meng-input kriteria yang diinginkan pembeli. Pemberian kriteria-kriteria dalam pemilihan Mobil bekas khususnya mobil dapat membantu dalam mengambil keputusan untuk menentukan mobil bekas sesuai keinginan dan kebutuhan konsumen.

Dengan menerapkan metode *K-Means Clustering* proses pemilihan mobil bekas lebih efisien sehingga konsumen lebih cepat memutuskan mobil yang diinginkan.

5. SARAN

Untuk pengembangan sistem selanjutnya, dapat diberikan saran-saran sebagai berikut :

1. Dikembangkan dalam aplikasi sistem khusus yang lebih efektif dan efisien.
2. Ditambahkan menu yang sesuai kebutuhan.
3. Data yang digunakan berasal dari negara yang akan disesuaikan dengan pengguna.

DAFTAR PUSTAKA

- [1] Abdillah, G., Putra, F. A., & Renaldi, F. (2016). Penerapan Data Mining Pemakaian Air Pelanggan untuk Menentukan Klasifikasi Potensi Pemakaian Air Pelanggan Baru di PDAM Tirta Raharja Menggunakan Algoritma K-Means. *Seminar Nasional Teknologi Informasi dan Komunikasi 2016*, 498-506.
- [2] Amirullah. (2015). *Metode Penelitian Manajemen*. Malang: Bayumedia Publishing Malang.
- [3] Leka, O. (2017). *Used Cars Database*. Retrieved November 27, 2017, from Kaggle: <http://www.kaggle.com/orgesleka/used-cars-database>
- [4] Nasari, F., & Darma, S. (2015). Penerapan K-Means Clustering pada Data Penerimaan Mahasiswa Baru (Studi Kasus : Universitas

Potensi Utama). *Seminar Nasional Teknologi Informasi dan Multimedia 2015*, 73-78.

- [5] Ong', J. O. (2013). Implementasi Algoritma K-Means Clustering untuk Menentukan Strategi Marketing President University. *Jurnal Ilmiah Teknik Industri*, 10-20.
- [6] Poerwanto, B., & Fa'rifah, R. Y. (2016). Analisis Cluster K-Means dalam Pengelompokkan Kemampuan Mahasiswa. *Jurnal Scientific Pinisi*, 92-96.
- [7] Pressman, R. S. (2001). *Software Engineering a Practitioner's Approach*. Thomas Casson.
- [8] Ridlo, M. R., Defiyanti, S., & Primajaya, A. (2017). Implementasi Algoritma K-Means Untuk Pemetaan Produktivitas Panen Padi di Kabupaten Karawang. *CITEE 2017*, 426-433.
- [9] Sumanti, T. (2013). Sistem Pendukung Keputusan Pemilihan Mobil Bekas dengan Menggunakan Metode Simple Additive Weighting (SAW). *Pelita Informatika Budi Darma*, 139-142.