

KLUSTERING JUMLAH PENDUDUK KOTA BANDUNG PER KELURAHAN DAN KECAMATAN BERDASARKAN STATUS PERKAWINAN PADA TAHUN 2019 DENGAN METODE K-MEANS

Jayanta¹, Fahira Hafizh Sekarani², Salsabilah Khansa³, Sarah
Ashari⁴

Fakultas Ilmu Komputer

Universitas Pembangunan Nasional Veteran Jakarta

Jl. Rs. Fatmawati, Pondok Labu, Jakarta Selatan, DKI Jakarta, 12450, Indonesia
anta.jayanta@gmail.com¹, fahirahafizh@gmail.com², salsabilahkhansa@ymail.com³,
azharsarah250@gmail.com⁴

Abstrak. Seiring bertambahnya dimensi waktu, penduduk pada setiap daerah di Indonesia semakin bertambah. Hal ini dikarenakan, tingginya angka kelahiran dan adanya transmigrasi. Penduduk yang melakukan transmigrasi tersebut memiliki harapan dengan mencari nafkah di kota besar akan merubah nasibnya. Selain faktor mata pencarian rezeki, penentuan status perkawinan di suatu daerah di Indonesia juga menjadi nilai komponen yang berguna untuk menjadi perhitungan tolak ukur dalam menentukan kepadatan di suatu daerah, seperti di Kota Bandung yang jumlahnya akan semakin bertambah sejalan dengan peningkatan frekuensi jumlah penduduk. Maksud dalam penulisan penelitian ini berguna untuk mengolah data penduduk Kota Bandung pada tahun 2019 yang diklasifikasikan berdasarkan status perkawinan dan yang dibagi per kelurahan dan kecamatan di Kota Bandung dengan metode klustering K-Means. Tujuan dilakukan klustering dimaksudkan untuk mencari padat atau tidaknya suatu kelurahan dan kecamatan di Kota Bandung berdasarkan status perkawinan.

Kata Kunci: K-Means, Klustering, Penduduk.

1 Pendahuluan

Fenomena peningkatan jumlah penduduk suatu daerah di Indonesia terjadi karena terkonsentrasinya suatu aktivitas dan kegiatan pada daerah tertentu sehingga, mengakitnya kepadatan di suatu daerah. Frekuensi intensitas penduduk yang melonjak untuk segelintir kelompok dan sektor membawa hal yang baik karena, jumlah penduduk semakin bertambah dapat menjadikan suatu hal yang menguntungkan dari sektor pembangunan, ekonomi, dan bertambahnya tenaga kerja di daerah tersebut. Peningkatan jumlah penduduk apabila tidak diseimbangi dengan faktor kualitas akal manusia seiring berjalannya waktu akan menjadi suatu permasalahan.

Data yang dipergunakan dalam penelitian ini diklasifikasikan berdasarkan status perkawinan suatu daerah berdasarkan daftar nama data kelurahan di Kota Bandung tahun 2019. Tujuan dilakukannya klustering dengan berdasarkan klasifikasi status perkawinan untuk membantu pemerintah dalam melakukan pemerataan pembangunan di Kota Bandung dalam melihat padat atau tidak padatnya suatu wilayah kelurahan dan kecamatan di Kota Bandung. Pemanfaatan teknologi data mining dapat membantu mempermudah pekerjaan dalam melakukan kegiatan klustering terhadap padat atau tidaknya penduduk yang mendiami suatu daerah kelurahan dan kecamatan di Kota Bandung.

Algoritma K-Means merupakan algoritma yang dilakukan dengan cara iterasi. Proses awal perhitungan algoritma dilakukan dengan cara menentukan objek K secara random yang ditentukan juga sesuai dengan berapa hasil kluster yang ingin dihasilkan diinisialisasikan dengan huruf "C" atau *centroid*. K-Means dipergunakan untuk membantu mengklusterkan padat atau tidak padatnya suatu kelurahan dan kecamatan di Kota Bandung berdasarkan status perkawinan. Inisialisasi kluster terbagi menjadi dua, yaitu C1 dan C2. Data penduduk tersebut adalah merupakan hasil pengumpulan data yang dipublikasikan melalui website bank data Kota Bandung yang terdiri atas dua jenis variabel

klasifikasi berupa status perkawinan (kawin dan belum kawin) sebanyak 152 kelurahan dan kecamatan yang berada di Kota Bandung.

2 Metodologi Penelitian

2.1 Data Mining

Pengertian dari data mining adalah suatu analisis terhadap data dengan ukuran besar yang tujuannya untuk mendapatkan pola penting pada suatu data (Sitanggang, 2014). Data mining pada dasarnya disebut juga dengan Knowledge-Discovery in Database (KDD). KDD merupakan serangkaian proses atau aktivitas secara otomatis dari pencarian data yang besar di sebuah memori yang nantinya didapatkan pola dengan penggunaan alat seperti pengelompokan (clustering), klasifikasi atau hubungan (association).

2.2 Clustering

Menurut Widodo (2013:9) Clustering atau klasifikasi adalah sebuah metode yang pemanfaatannya untuk membagi rangkaian data dari kesamaan yang sudah ditentukan menjadi beberapa kelompok. Pengertian dari cluster adalah beberapa objek data yang berkumpul dengan kemiripan antara data satu dengan data lainnya dalam cluster yang sama dan berbeda dengan objek yang clusteranya berbeda. Nantinya objek dikelompokkan ke satu atau lebih cluster yang mempunyai kesamaan yang paling tinggi antara satu dengan lainnya sehingga objek yang sama akan terkumpul dalam satu cluster. Metode clustering bersifat unsupervised (tanpa arahan), maksudnya adalah metode ini tidak membutuhkan target untuk outputnya dan proses yang dilakukan tanpa latihan (training).

2.3 K-Means

Algoritma K-Means merupakan metode non-hierarchical dimana metode ini memisahkan suatu data menjadi satu atau lebih cluster yang nantinya dalam satu cluster data tersebut memiliki karakteristik yang sama dan data dengan karakteristik berbeda dikelompokkan ke dalam cluster lain (Rahman, 2017). Metode penelitian yang digunakan adalah metode K-Means dengan algoritma adalah sebagai berikut (Elisawati, Wahyuni and Arianto, 2019).

1. Menentukan k sebagai jumlah cluster yang ingin dibuat
2. Membuat nilai *random* untuk pusat *cluster (centroid)* awal sebanyak k.
3. Menghitung jarak antar data terhadap masing-masing centroid menggunakan rumus *euclidian distance* yang nantinya didapat jarak yang paling dekat dari setiap data dengan centroid. Persamaan yang digunakan untuk menghitung *euclidian distance* adalah

$$d_i = \sqrt{\sum_{j=1}^n (x_{ij} - c_{kj})^2}$$

Dimana :

$$d_i = \sqrt{\sum_{j=1}^n (x_{ij} - c_{kj})^2} \quad (1)$$

4. Mengklasifikasi setiap data terhadap kedekatannya dengan centroid (jarak terkecil)

$$\text{Min} \sum_{j=1}^k \sum_{i=1}^n d_{ij} = 1 =$$

$$\sqrt{\sum_{j=1}^k \sum_{i=1}^n (x_{ij} - c_{kj})^2} \quad (2)$$

atau = MIN (C_{ij} : C_{kj} : C_{lj})

5. Nilai *centroid* diperbaharui. Nilai *centroid* didapat dari rata-rata cluster yang bersangkutan

$$C_{kj} =$$

$$\sum_{k=1}^K \frac{1}{n_k} = 1$$

$$\diamond (3)$$

6. Perulangan dari langkah 2 hingga langkah kelima sampai anggota cluster tidak ada yang berubah.

7. Jika langkah keenam sudah terpenuhi, maka nilai pusat cluster pada iterasi terakhir akan digunakan sebagai parameter untuk menentukan klasifikasi data.

Dibawah ini terdapat flowchart yang menjelaskan langkah penerapan metode K-Means yang ditampilkan pada Gambar 1.

1.



Gambar 1. Flowchart K-Means Clustering

2.4 Datasets

Perolehan data didapatkan dari data jumlah penduduk berdasarkan kategori belum kawin dan kawin berdasarkan daftar

nama kelurahan di Kota Bandung pada tahun 2019. Data ini berasal dari situs *website* Data Kota Bandung yang dikumpulkan pada tahun 2019 dapat dilihat pada Tabel 1 berikut.

Tabel 1. Data Kependudukan di Kota Bandung per Kelurahan

Kecamatan	Kelurahan Belum Kawin	Kawin	Kecamatan	Kelurahan Belum Kawin	Kawin
Andir	Campaka	8310	Arcaman	Cisaranten Bina	5389 5938
Andir	Ciroyom	9279	Andir	Harapan	9439
Andir	Dungus Cariang	8522	Arcaman	Cisaranten Kulon	10377 71147
Andir	Garuda	4852	Arcaman	Sindanglaya/Cisaranten Endah	8995 9634
Andir	Kebon Jeruk	5361	Arcaman	Sukamiskin	9638 1066
Andir	Maleber	9016	Andir	Maleber	9915
Antapani	Antapani Kidul	11390	Antapani	Antapani Kidul	12581
Antapani	Antapani Kulon	4884	Antapani	Antapani Kulon	5479
Antapani	Antapani Tengah	10335	Antapani	Antapani Tengah	11906
Antapani	Antapani Wetan	8516	Antapani	Antapani Wetan	9221
			Astana Anyar	Cibadak	6172 6222
			Astana Anyar	Karanganyar	3060 3194
			Astana Anyar	Karasak	6195 6565

Kecamat	Kelurahan	Belum Kawin	Kawin	Kecamat	Kelurahan	Belum Kawin	Kawin
Astana	Nyengseret	5686	5739	Babakan	Cirangrang	5069	5637
Anyar				Ciparay			
Astana	Panjunan	3557	3567	Babakan	Margahayu	10030	10961
Anyar				Ciparay	Utara		
Astana	Pelindung	9518	9651	Babakan	Margasuka	5347	5873
Anyar	Hewan			Ciparay			
Babakan	Babakan	13861	14754
Ciparay			
Babakan	Babakan	14934	15954	Ujung	Pasirwangi	7106	8203
Ciparay	Ciparay			Berung			

3 Hasil dan Pembahasan

3.1 Clustering

Penelitian ini menggunakan data yang terdiri dari 151 record. Data ini memiliki 2 variable yaitu variable belum kawin dan variable kawin. Skenario dalam penelitian ini memprediksi bertambahnya jumlah penduduk berdasarkan jumlah perkawinan pada suatu wilayah. Nantinya, wilayah ini akan diklasterisasi ke dalam klaster yang jumlah penduduknya cenderung akan bertambah atau klaster yang jumlah penduduknya cenderung akan tetap atau bertambahnya tidak signifikan.

Dalam melakukan klasterisasi, penelitian ini menggunakan K-Means Clustering yang perlu beberapa iterasi sampai menemukan centroid stabil. Dalam penelitian ini, terdapat dua klaster, centroid dipilih secara acak, dan iterasi dilakukan sampai lima kali. Dijelaskan dalam tabel di bawah ini,

Tabel 2 Iterasi dan Nilai Centroid

Iterasi	Centroid 1		Centroid 2	
	Variabel BK	Variabel K	Variabel BK	Variabel K
1.	8310	8976	9279	9439
2.	5645,970297	6007,33663	11117,28	11953,48
3.	7206,340659	7659,054945	7838,833333	8457,35
4.	4795,202532	5086,063291	10378,97222	11147,44444
5.	4861,740741	5160,888889	10461,51429	11234,04286
6.	4861,740741	5160,888889	10461,51429	11234,04286

Tabel 3 Hasil Iterasi Ke-5

No. Kls	Kecamatan	Kelurahan		No.	Kecamatan	Kelurahan	Kls
1	Andir	Campaka	C2	4	Andir	Garuda	C1
2	Andir	Ciroyom	C2	5	Andir	Kebon Jeruk	C1
	Andir	Dungus Cariang	C2	6	Andir	Maleber	C2

3

No.	Kecamatan	Kelurahan	Kls	No.	Kecamatan	Kelurahan	Kls
7	Antapani	Antapani Kidul	C2	26	Babakan Ciparay	Sukahaji	C2
8	Antapani	Antapani Kulon	C1	27	Bandung Kidul	Batununggal	C2
9	Antapani	Antapani Tengah	C2	28	Bandung Kidul	Kujangsari	C2
10	Antapani	Antapani Wetan	C2	29	Bandung Kidul	Mengger	C1
11	Arcamanik	Cisaranten Bina Harapan	C1	30	Bandung Kidul	Wates	C1
12	Arcamanik	Cisaranten Kulon	C2	31	Bandung Kulon	Caringin	C1
13	Arcamanik	Sindanglaya/Cis antren Endah	C2	32	Bandung Kulon	Cibuntu	C2
14	Arcamanik	Sukamiskin	C2	33	Bandung Kulon	Cigondewah Kaler	C2
15	Astana Anyar	Cibadak	C1	34	Bandung Kulon	Cigondewah Kidul	C1
16	Astana Anyar	Karanganyar	C1	35	Bandung Kulon	Cigondewah Rahayu	C1
17	Astana Anyar	Karasak	C1	36	Bandung Kulon	Cijerah	C2
18	Astana Anyar	Nyengseret	C1	37	Bandung Kulon	Gempolsari	C2
19	Astana Anyar	Panjunan	C1	38	Bandung Kulon	Warung Muncang	C2
20	Astana Anyar	Pelindung Hewan	C2	39	Bandung Wetan	Cihapit	C1
21	Babakan Ciparay	Babakan	C2	40	Bandung Wetan	Citarum	C1
22	Babakan Ciparay	Babakan Ciparay	C2	41	Bandung Wetan	Tamansari	C2
23	Babakan Ciparay	Cirangrang	C1	42	Batununggal	Binong	C2
24	Babakan Ciparay	Margahayu	C2	43	Batununggal	Cibangkong	C2
25	Batununggal	Maleara	C2	44	Batununggal	Gumuruh	C2
26	Batununggal	Samoja	C1	45	Batununggal	Kacapiring	C1
27	Babakan Ciparay	Margasuka	C1	46	Gibeunying Batununggal Kidul	Cikutra Kebongedang	C2
28	Bojongloa Kaler	Babakan Asih	C1	47	Gibeunying Batununggal Kidul	Padasuka Kebonwaru	C2
29	Bojongloa Kaler	Babakan Tarogong	C2	48	Cibeunying Kidul	Pasirlayung	C2

52	Bojongloa Kaler	Jamika	C2
53	Bojongloa Kaler	Kopo	C2
54	Bojongloa Kaler	Suka Asih	C2
55	Bojongloa Kidul	Cibaduyut	C1
56	Bojongloa Kidul	Cibaduyut Kidul	C1
57	Bojongloa Kidul	Cibaduyut Wetan	C1
58	Bojongloa Kidul	Kebon Lega	C2
59	Bojongloa Kidul	Mekarwangi	C1
60	Bojongloa Kidul	Situsaeur	C2
61	Buahbatu	Cijaura	C2
62	Buahbatu	Jatisari	C1
63	Buahbatu	Margasari	C2
64	Buahbatu	Sekejati	C2
65	Cibeunying Kaler	Cigadung	C2
66	Cibeunying Kaler	Cihaur Geulis	C1
67	Cibeunying Kaler	Neglasari	C1
68	Cibeunying Kaler	Sukaluyu	C1
69	Cibeunying Kidul	Cicadas	C1

No.	Kecamatan	Kelurahan	Kls
99	Gedebage	Cisantren Kidul	C2
100	Gedebage	Rancabolang	C1
101	Gedebage	Rancanumpang	C1

 104
 102

 105
 103

73	Cibeunying Kidul	Sukamaju	C1
74	Cibeunying Kidul	Sukapada	C2
75	Cibiru	Cipadung	C2
76	Cibiru	Cisurupan	C1
77	Cibiru	Palasari	C2
78	Cibiru	Pasir Biru	C1
79	Cicendo	Arjuna	C1
80	Cicendo	Husen Sastranegara	C1
81	Cicendo	Pajajaran	C2
82	Cicendo	Pamoyanan	C1
83	Cicendo	Pasir Kaliki	C1
84	Cicendo	Sukaraja	C2
85	Cidadap	Ciumbuleuit	C2
86	Cidadap	Hegarmanah	C2
87	Cidadap	Ledeng	C1
88	Cinambo	Babakan Penghulu	C1
89	Cinambo	Cisaranten Wetan	C1
90	Cinambo	Pakemitan	C1
91	Cinambo	Sukamulya	C1
92	Coblong	Cipaganti	C1
93	Coblong	Dago	C2
94	Coblong	Lebak Siliwangi	C1
95	Coblong	Lebakgede	C1
96	Coblong	Sadangserang	C2
97	Coblong	Sekeloa	C2
98	Gedebage	Cimenerang	C1

No.	Kecamatan	Kelurahan	Kls
123	Rancasari	Cipamokolan	C2
124	Rancasari	Derwati	C2
125	Rancasari	Mekarmulya	C2
126	Rancasari	Munjahlega	C2

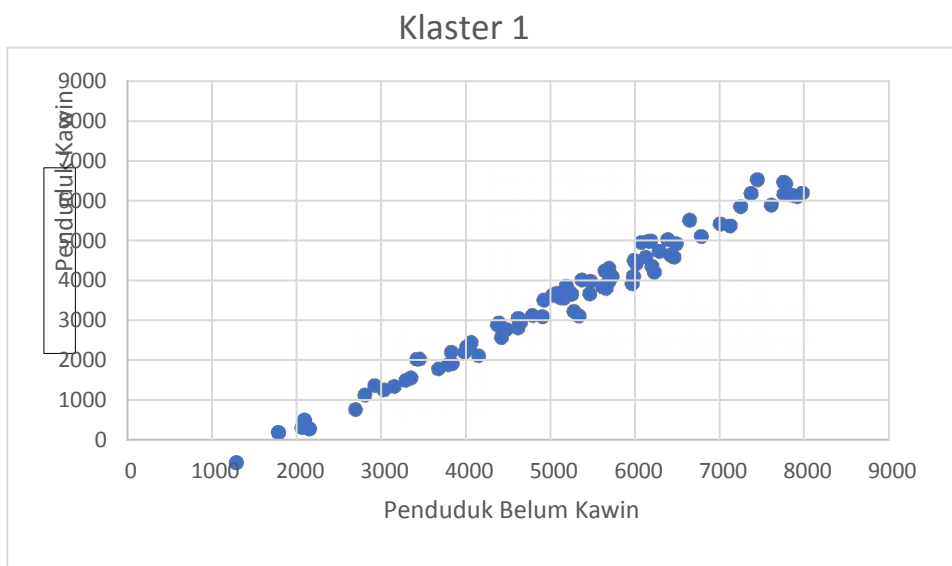
106

 107

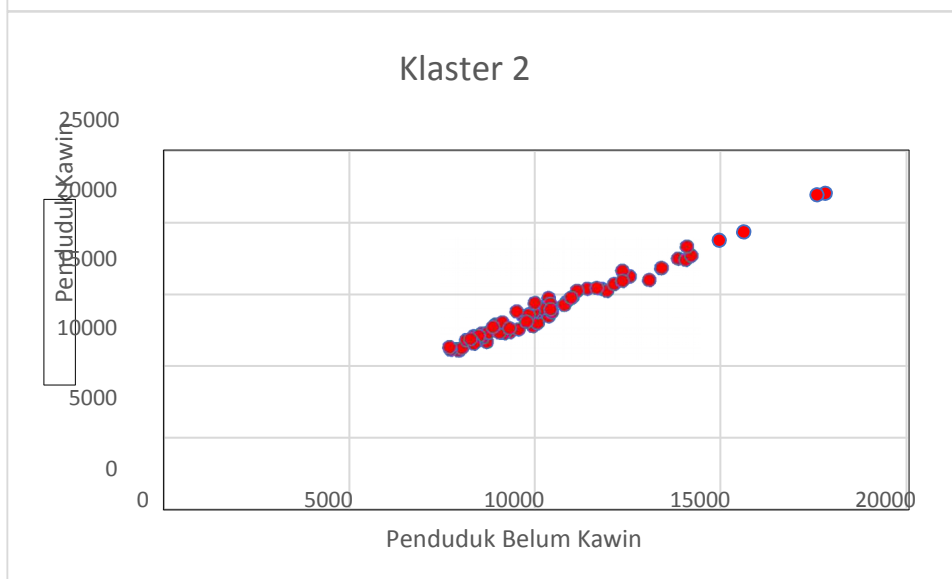
	Kiaracondong	Cicaheum	C2
Kiaracondong	Kiaracondong	Kebon Jayanti	C1
Babakan C2	Kiaracondong	Kebon Kangkung	C1
	S ^S Kiaracondong	Sukapura	C2
Kiaracondong Babakansari C2			
108	Lengkong	Burangrang	C1
109	Lengkong	Cijagra	C1
110	Lengkong	Cikawao	C1
111	Lengkong	Lingkar Selatan	C1
112	Lengkong	Malabar	C1
113	Lengkong	Paledang	C1
114	Lengkong	Turangga	C1
115	Mandalajati	Jatihandap	C2
116	Mandalajati	Karang Pamulang	C2
117	Mandalajati	Pasir Impun	C1
118	Mandalajati	Sindang Jaya	C1
119	Panyileukan	Cipadung Kidul	C1
120	Panyileukan	Cipadung Kulon	C1
121	Panyileukan	Cipadung Wetan	C1
122	Panyileukan	Mekar Mulya	C1

127	Regol	Ancol	C1
128	Regol	Balonggede	C1
129	Regol	Ciateul	C1
130	Regol	Cigereleng	C1
131	Regol	Ciseureuh	C1
132	Regol	Pasirluyu	C2
133	Regol	Pungkur	C1
134	Sukajadi	Cipedes	C2
135	Sukajadi	Pasteur	C2
136	Sukajadi	Sukabungah	C2
137	Sukajadi	Sukagalih	C2
138	Sukajadi	Sukawarna	C1
139	Sukasari	Gegerkalong	C2
140	Sukasari	Isola	C1
141	Sukasari	Sarijadi	C2
142	Sukasari	Sukarasa	C1
143	Sumur Bandung	Babakan Ciamis	C1
144	Sumur Bandung	Braga	C1
145	Sumur Bandung	Kebon Pisang	C1
146	Sumur Bandung	Merdeka	C1
147	Ujung Berung	Cigending	C1
148	Ujung Berung	Pasanggrahan	C2
149	Ujung Berung	Pasir Endah	C1
150	Ujung Berung	Pasirjati	C2
151	Ujung Berung	Pasirwangi	C1

Dari hasil klasterisasi di atas mendapatkan hasil bahwa 81 wilayah masuk ke dalam centroid 1 dan 70 wilayah yang masuk ke dalam centroid 2. Centroid 1 mewakili kluster yang diprediksi jumlah penduduknya cenderung akan bertambah atau penambahannya signifikan. Sementara centroid 2 mewakili kluster yang diprediksi jumlah penduduknya cenderung akan tetap atau bertambahnya tidak signifikan. Record yang termasuk ke dalam kluster 1 dan kluster 2 ditampilkan dalam diagram sebagai berikut,



Gambar 2. Chart Hasil Iterasi Terakhir (Klaster 1)



Gambar 3. Chart Hasil Iterasi Terakhir (Klaster 2)

4 Kesimpulan

Hasil penelitian klusterisasi dengan menggunakan dataset dapat ditarik kesimpulan bahwa algoritma K-Means melakukan klusterisasi terhadap data yang telah digunakan. Pengolahan data telah dibagi menjadi dua cluster, yaitu C1 dan C2 di mana cluster 1 mewakili wilayah yang tidak terjadi peningkatan yang signifikan sedangkan, cluster 2 mewakili wilayah yang terjadi peningkatan penduduk yang cukup signifikan. Jumlah data clustering dari C1 sebanyak

81 data wilayah sedangkan, C2 sebanyak 70 wilayah.

Berdasarkan penelitian yang sudah dilakukan maka saran bagi penelitian yang akan datang dapat menggunakan algoritma clustering yang lain yang di mana besar kemungkinan akan menghasilkan hasil yang baik.

Referensi

- [1] Elisawati, Wahyuni, D. and Arianto, A. (2019) 'Analisa Clustering Pada Data Pelanggaran Lalulintas Di Pengadilan Negeri Dumai Dengan Menggunakan Metode K-Means', *Jiska*, 3(3), pp. 50–61. doi: <http://dx.doi.org/10.14421/jiska.2019.%25x>.
- [2] Rahman, A. T. (2017) 'Coal Trade Data Clustering Using K-Means (Case Study Pt. Global Bangkit Utama)', *ITSMART: Jurnal Teknologi dan Informasi*, 6(1), pp. 24–31. doi: 10.20961/ITS.V6I1.11296.
- [3] Sitanggang, Lenny Fitriany. 2014. Aplikasi Data Mining Association Rule dengan Algoritma FP – Growth untuk Mengenali Pola Pembelian Pelanggan pada Toko Kue (Studi Kasus: L'cheese Factory). Politeknik Caltex Riau: Pekanbaru.
- [4] Widodo. 2004. Psikologi Belajar. Jakarta: Rineka Cipta.