
Algoritma K-Means Untuk *Clustering* Rute Perjalanan Wisata Pada Agen Tour & Travel**Eni Irfiani, Fintri Indriyani**

Eni.enf@bsi.ac.id, fintri.fni@bsi.ac.id

Universitas Bina Sarana Informatika

Informasi Artikel

Diterima : 03-02-2020

Direview : 11-02-2020

Disetujui : 29-02-2020

Kata Kunci*clustering*; data mining;
algoritma K-Means; rute
perjalanan wisata

Abstrak

Dukungan pemerintah terhadap pengembangan pariwisata berdampak pada tumbuhnya peluang bisnis bagi agen perjalanan. Seiring dengan majunya sektor perjalanan domestik, bermunculan bentuk usaha agen *tour & travel* sehingga berpengaruh terhadap persaingan bisnis antara agen perjalanan. Permasalahan yang terdapat pada agen *tour & travel* yaitu kurangnya informasi tentang rute perjalanan wisata yang paling diminati pelanggan. Untuk menyelesaikan masalah tersebut metode yang digunakan untuk mengelompokkan rute perjalanan yang paling diminati dengan menggunakan metode dari data mining yaitu *clustering* dengan algoritma K-Means. Berdasarkan hasil penelitian didapatkan tiga kelompok rute perjalanan yaitu rute perjalanan wisata paling diminati sebesar 20%, perjalanan yang sedang diminati sebesar 30% dan perjalanan yang kurang diminati sebesar 50%.

Keywords*clustering*; data mining; K-
Means Algorithm; tourist
travel routes

Abstract

Government support for the development of tourism has an impact on the growth of business opportunities for travel agents. Along with the advancement of the domestic travel sector, tour & travel agent business forms have sprung up that influence business competition between travel agents. The problem with tour & travel agents is the lack of information about tourist routes that are most in-demand by customers. To solve this problem the method used to classify the most desirable travel routes using the method of data mining is clustering with the K-Means algorithm. Based on the results of the study found three groups of travel routes, namely the most desirable travel routes by 20%, the trips that are in demand by 30% and less desirable trips by 50%.

A. Pendahuluan

Teknologi informasi memiliki pengaruh besar bagi industri pariwisata. Berdasarkan data biro pusat statistik dalam kurun waktu dari tahun 2013 sampai tahun 2018 terdapat peningkatan jumlah perjalanan (*trips*) wisatawan nusantara sebesar 21,34 persen. Jumlah wisatawan nusantara yang melakukan kunjungan pada tahun 2018 sebesar 303,4 juta perjalanan [1]. Tingginya angka perjalanan wisatawan nusantara, mampu menghidupkan perekonomian pada sektor angkutan, industri lokal, perhotelan dan lain-lain. Peningkatan permintaan wisatawan terhadap jasa angkutan mendorong perusahaan untuk terus melakukan inovasi guna mengembangkan usaha jasa tour & travel dengan menggunakan teknologi modern. Permasalahan yang terdapat pada perusahaan yaitu kurangnya informasi mengenai pengelompokan rute perjalanan wisata yang paling diminati oleh wisatawan nusantara khususnya yang berada di daerah Tangerang dan sekitarnya.

Kesuksesan bisnis pada bidang pariwisata sangat bergantung pada penggunaan data sehari-hari. Data mampu menjadi kunci sukses keberhasilan suatu bisnis. Untuk menentukan paket wisata di masa yang akan datang diperlukan data pesanan paket wisata yang terjual melalui reservasi *online*. Berdasarkan hasil penelitian, dengan metode *k-means* data tersebut dikelompokkan berdasarkan produk, negara dan lokasi tujuan [2].

Wisata kuliner merupakan suatu kegiatan yang sering dilakukan wisatawan. Kendala pada saat melakukan wisata kuliner yaitu kurangnya informasi terhadap lokasi yang menyediakan wisata kuliner. Untuk membantu para wisatawan dalam menentukan lokasi dari wisata kuliner yang mempunyai fasilitas lengkap maka perlu adanya metode pengelompokan data sesuai dengan *cluster* masing-masing. Dengan metode *k-means* data dikelompokkan berdasarkan jarak yang dihitung dari posisi awal wisatawan ke alamat lokasi wisata tujuan [3].

Data *mining* dapat dideskripsikan sebagai suatu metode untuk menganalisis data dalam jumlah besar untuk menemukan *prototipe*, model dan hubungan dari data yang sebelumnya tidak diketahui untuk nantinya sebagai informasi dalam pengambilan keputusan. Kondisi persaingan bisnis pada jasa perjalanan dan pariwisata sangat ketat, untuk itu perlu adanya metode data *mining* dalam mengembangkan produk dan layanan pariwisata [4].

Setiap tahun wisatawan datang dari berbagai tempat dari seluruh dunia. Industri pariwisata menjadi sumber perekonomian nasional dan membuka banyak lapangan pekerjaan bagi masyarakat sekitar. Perusahaan pariwisata harus lebih efektif dalam memperkaya fasilitas untuk mengetahui pelanggan yang sering datang. Oleh sebab itu perlu ada penelitian yang fokus terhadap pengembangan industri pariwisata. Dengan menggunakan teknik data *mining*, data akan dikumpulkan dari perusahaan pariwisata dan hasilnya disajikan dalam bentuk grafik serta informasi sehingga membantu manajemen perusahaan untuk merubah bisnis menjadi lebih baik [5].

Yogyakarta merupakan kota yang banyak dikunjungi oleh wisatawan domestik dan Internasional. Salah satu akun instagram @explorejogya memberikan informasi tujuan wisata yang dapat dikunjungi di Yogyakarta. Oleh karena itu diperlukan metode *k-means* untuk mengklasifikasi popularitas tujuan wisata berdasarkan tempat yang paling tinggi diminati, sedang diminati dan kurang diminati [6]

Perkembangan jasa akomodasi pariwisata di Indonesia semakin menunjukkan perkembangan pesat. Pemerintah daerah sedang menggiatkan ekonomi kreatif di berbagai bidang termasuk daya tarik wisata, namun kurangnya informasi serta pengelompokan kunjungan wisatawan dari berbagai obyek wisata menyebabkan tempat wisat yang kurang dikenal jarang mendapatkan kunjungan dari wisatawan [7].

Untuk mengetahui informasi tersembunyi perlu dilakukan pengolahan data. Dibutuhkan strategi khusus pada divisi marketing untuk melakukan pemasaran agar promosi menjadi lebih efektif dan efisien. Metode yang digunakan untuk mengolah data tersembunyi menggunakan *k-means clustering*. Melalui metode ini data akan dikelompokkan dalam beberapa *cluster* berdasarkan kemiripan [8].

Algoritma *k-means* merupakan suatu metode yang tidak dipengaruhi oleh runtutan objek yang akan digunakan namun dalam penentuan titik awal pusat *cluster* dapat dilakukan secara acak dari salah satu objek pada saat dilakukan perhitungan [9].

Industri pariwisata sangat membutuhkan penggunaan teknologi informasi guna meningkatkan efisiensi operasional perusahaan serta meningkatkan kualitas pelayanan kepada pelanggan. Agen wisata umumnya menggunakan layanan internet sebagai saluran komunikasi dengan pelanggan dikarenakan dapat mudah diakses dan murah. Perilaku wisatawan cenderung berubah seiring berkembangnya penggunaan aplikasi pada *smartphone*. Wisatawan menghindari membuat rencana liburan karena yang diinginkan adalah liburan dengan pengalaman baru. Pada sektor pariwisata memahami kebutuhan pelanggan merupakan suatu keuntungan sehingga dapat merespon dengan cepat dengan penawaran-penawaran menarik. Perusahaan yang bergerak di bidang pariwisata harus mampu mendefinisikan strategi pemasaran dan menyediakan produk dan layanan yang terjangkau bagi wisatawan. Penggunaan *data mining* sangat cocok untuk mengumpulkan informasi dan perilaku pelanggan untuk dipelajari. Algoritma *data mining* digunakan untuk menentukan pola dalam industri pariwisata. Data yang disajikan dalam jumlah banyak kemudian disaring untuk menemukan informasi yang berguna. Saat ini penggunaan algoritma *data mining* sangat populer pada aplikasi perjalanan dan pariwisata. Melalui algoritma *data mining* akan terbentuk aturan (*rule*) untuk menentukan pola perilaku atau keadaan tertentu [10] [11].

Salah satu metode penggalian data yang saat ini sedang diminati untuk digunakan pada penelitian yaitu *data mining*. Melalui teknik *data mining* penelitian menggunakan teknologi *database* yang dikelola secara dinamis. Guna mengekstrak data yang berjumlah besar diperlukan suatu cara mengeksplorasi agar lebih efisien. Proses penggalian data merubah data yang tidak berguna menjadi sesuatu yang besar manfaatnya [12].

Salah satu metode yang digunakan pada *data mining* untuk melakukan pengelompokan data yaitu *Clustering*. *Cluster* merupakan teknik pengumpulan objek yang memiliki kemiripan antara satu objek dengan objek lain. Sedangkan *clustering* atau analisis *cluster* merupakan metode pengelompokan satu set objek secara fisik atau abstrak ke dalam objek yang sama. *K-Means clustering* merupakan salah satu metode dalam *data mining*, yaitu proses pemodelan tanpa perlu dilakukan *supervise* dan merupakan metode pengelompokan data dilakukan secara partisi. Dimana setiap

kelompok memiliki karakteristik yang menyerupai namun mempunyai karakteristik yang berbeda dengan kelompok lainnya [13].

Berdasarkan penelitian sebelumnya guna mendapatkan informasi tentang rute perjalanan yang paling diminati pelanggan, perlu dilakukan pengelompokan data berdasarkan rute perjalanan wisata yang ada pada agen *tour & travel*. Metode yang digunakan untuk menyelesaikan pengelompokan data yang paling diminati yaitu metode data *mining* dengan *clustering* menggunakan algoritma *k-means*. Pengelompokan rute perjalanan wisata yang akan ditentukan antara lain rute perjalanan paling diminati, rute perjalanan sedang diminati dan rute perjalanan kurang diminati. Hasil dari *clustering* rute perjalanan akan didapatkan grafik serta persentase pengelompokan rute paling diminati, sedang diminati dan kurang diminati.

B. Metode Penelitian

Dalam penelitian ini tahapan penelitian yang dilakukan adalah [14]:

1. Survey Acuan
Mengumpulkan bahan acuan dan informasi rute perjalanan wisata yang berkaitan dengan penelitian.
2. Identifikasi Masalah
Mengidentifikasi permasalahan yang ada yaitu berdasarkan data rute perjalanan selama dua tahun selanjutnya akan memilih permasalahan berdasarkan kelompok atau *kluster* di tiap data tersebut.
3. Melakukan Kajian Pustaka
Mencari sumber referensi untuk digunakan sebagai kajian pustaka penelitian. Berdasarkan sumber referensi, data yang telah didapatkan akan diolah menggunakan metode *clustering* dan algoritma *k-means*.
4. Menentukan variabel dan sumber data
Menentukan variabel dari data rute perjalanan terlebih dahulu kemudian menentukan data mana yang akan di ambil sesuai kebutuhan penelitian.
5. Menyusun Penelitian
Pada tahap ini menentukan data dari agen perjalanan selama dua tahun yang akan digunakan untuk penelitian.
6. Mengumpulkan data
Meminta data rute perjalanan melalui staf agen perjalanan mulai dari tahun 2017 sampai 2018.
7. Mengolah Data
Data yang sudah ada kemudian diolah dan dihitung menggunakan algoritma *k-means* untuk di kelompokkan berdasarkan kemiripan karakteristik yaitu rute paling diminati, rute sedang diminati dan rute kurang diminati. Data akan dihitung secara manual serta menggunakan tools *Rapid Miner*.
8. Menganalisis Data
Menganalisis data yang sudah diolah berdasarkan hasil perhitungan secara manual ataupun hasil pengolahan data menggunakan tools *Rapid Miner*.
9. Menarik kesimpulan dan saran
Dari hasil analisis akan ditarik kesimpulan sesuai dengan permasalahan dan solusi yang akan diselesaikan serta saran untuk pengembangan penelitian berikutnya

Teknik pengumpulan data yang digunakan pada penelitian ini antara lain:

- a. Data kuantitatif
Mengumpulkan data-data rute perjalanan wisata selama dua tahun dari agen tour & travel yang terdapat di kota Bogor untuk menentukan permasalahan pada penelitian.
- b. Data Primer
Data diperoleh secara langsung dari PT. Nirwana Tour & Travel Tangerang.
- c. Data Sekunder
Data yang dikumpulkan berdasarkan dokumen pendukung lainnya seperti: nama pelanggan, tujuan wisata, tempat penjemputan dan jarak tempuh.
- d. Populasi dan sampel penelitian
Populasi merupakan wilayah generalisasi yang terdiri dari objek yang memiliki kualitas serta karakteristik yang telah ditetapkan oleh peneliti yang kemudian akan ditarik kesimpulan. Pada penelitian ini populasi yang diambil sebanyak 32 data rute perjalanan selama dua tahun, dari populasi tersebut diambil sebanyak 29 sampel.

Langkah-langkah pembentukan *clustering* dengan menggunakan K-Means antara lain [3]:

1. Menentukan jumlah *cluster* yang ingin dibentuk.
2. Menentukan nilai awal dari pusat *cluster* (*centroid*) sejumlah yang akan dibentuk.
3. Melakukan perhitungan jarak dari setiap data terhadap *centroid* menggunakan rumus Euclidian Distance.

$$d(X_i, \mu_j) = \sqrt{(x_i - \mu_j)^2} \quad (1)$$

4. Mengalokasikan setiap data berdasarkan jarak dengan *centroid* yang terkecil
5. Melakukan perubahan *centroid* baru dengan menghitung nilai rata-rata *cluster*
6. Kembali ke langkah ke-2 sampai langkah ke-5 sampai nilai pusat *cluster* tidak mengalami perubahan.

C. Hasil dan Pembahasan

Data rute perjalanan yang digunakan dalam penelitian sebagai rute tujuan wisata berdasarkan dua kriteria jumlah perjalanan dan jarak tempuh dapat dilihat pada Tabel 1.

Tabel 1. Data Rute Perjalanan

No.	Rute Perjalanan	Data Perjalanan	
		Jumlah Perjalanan (JP)	Jarak Tempuh (JT)
1	Tangerang - Pantai Anyer	14	114
2	Tangerang - Bandung	8	188
3	Tangerang - Jogja	10	598
4	Tangerang - Puncak	6	111
5	Tangerang - Kebun Binatang Ragunan	4	43
6	Tangerang - Bromo	3	888

7	Tangerang - Banten	3	73
8	Tangerang - Pulo Saiji Cisauk	2	12
9	Tangerang - Tangkuban Perahu	2	197
10	Tangerang - Cirebon	1	255
11	Tangerang - Pantai Sawarna	2	169
12	Tangerang - Pamijahan	1	68
13	Tangerang - Museum Fatahillah	3	56
14	Tangerang - The Jungle Bogor	2	58
15	Tangerang - Pantai Ancol	2	49
16	Tangerang - Malang	4	888
17	Tangerang - Batu Raden	2	402
18	Tangerang - Kawah Putih	2	223
19	Tangerang - CiKundul	2	120
20	Tangerang - Pantai Ujung Genteng	1	238
21	Tangerang - Kediri	1	747
22	Tangerang - Pantai Santolo	1	318
23	Tangerang - Pelabuhan Ratu	1	136
24	Tangerang - Kuningan	1	41
25	Tangerang - Sumedang	1	239
26	Tangerang - Pulau Seribu	1	40
27	Tangerang - Tebing Keraton	1	195
28	Tangerang - Gucci Tegal	1	365
29	Tangerang - Blitar	1	787

Tahapan yang digunakan untuk mengklasifikasi menggunakan algoritma *k-means* adalah menentukan *cluster* yang akan digunakan, membagi *cluster* menjadi tiga *cluster* yaitu: rute perjalanan paling diminati (C0), rute perjalanan sedang diminati (C1) dan rute perjalanan kurang diminati (C2). Menentukan nilai *centroid* (awal) yang baru, didapat nilai *centroid* baru yang terdapat pada Tabel 2.

Tabel 2. Nilai *Centroid* Baru

	C0	C1	C2
Jumlah Perjalanan	0,744	0,096	0,07
Jarak Tempuh	0,329	0,931	0,163

Langkah selanjutnya adalah menentukan jarak *cluster* terdekat menggunakan rumus (1), hasil perhitungan jarak *cluster* terdekat dapat dilihat pada Tabel 3.

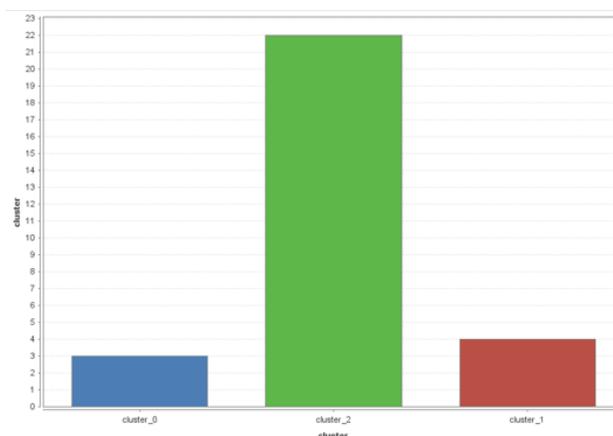
Tabel 3. Data Jarak *Cluster* Terdekat

No.	Rute Perjalanan	Jarak C0	Jarak C1	Jarak C2
1	Tangerang - Pantai Anyer	0,333	1,217	0,931
2	Tangerang - Bandung	0,242	0,854	0,47
3	Tangerang - Jogja	0,344	0,651	0,802
4	Tangerang - Puncak	0,419	0,867	0,319
5	Tangerang - Kebun Binatang Ragunan	0,591	0,906	0,205
6	Tangerang - Bromo	0,894	0,09	0,841

7	Tangerang - Banten	0,645	0,863	0,125
8	Tangerang - Pulo Saiji Cisauk	0,744	0,931	0,163
9	Tangerang - Tangkuban Perahu	0,677	0,72	0,049
10	Tangerang - Cirebon	0,746	0,661	0,134
11	Tangerang - Pantai Sawarna	0,684	0,752	0,018
12	Tangerang - Pamijahan	0,79	0,872	0,121
13	Tangerang - Museum Fatahillah	0,653	0,883	0,141
14	Tangerang - The Jungle Bogor	0,722	0,879	0,111
15	Tangerang - Pantai Ancol	0,726	0,889	0,121
16	Tangerang - Malang	0,845	0,151	0,852
17	Tangerang - Batu Raden	0,677	0,486	0,282
18	Tangerang - Kawah Putih	0,673	0,69	0,078
19	Tangerang - CiKundul	0,698	0,808	0,04
20	Tangerang - Pantai Ujung Genteng	0,747	0,68	0,118
21	Tangerang - Kediri	0,902	0,133	0,68
22	Tangerang - Pantai Santolo	0,744	0,59	0,199
23	Tangerang - Pelabuhan Ratu	0,767	0,795	0,073
24	Tangerang - Kuningan	0,801	0,903	0,148
25	Tangerang - Sumedang	0,747	0,679	0,119
26	Tangerang - Pulau Seribu	0,801	0,904	0,149
27	Tangerang - Tebing Keraton	0,754	0,728	0,084
28	Tangerang - Gucci Tegal	0,748	0,537	0,25
29	Tangerang - Blitar	0,929	0,107	0,725

Berdasarkan Tabel 2. dihitung kembali nilai *centroid* baru yang hasil perhitungan nilai *centroid* baru memiliki kesamaan nilai dengan nilai sebelumnya. Proses perhitungan dihentikan dikarenakan hasil telah sama maka proses perhitungan nilai *centroid* berakhir seperti yang tertera pada Tabel 3.

Berdasarkan hasil akhir nilai *cluster* terdekat didapat hasil pengelompokkan sebagai berikut: *cluster* C0 merupakan rute paling diminati dengan jumlah sebanyak 20%, C1 merupakan rute yang sedang diminati sebanyak 30%, sedangkan C2 merupakan rute yang kurang diminati sebanyak 50%. Selanjutnya hasil perhitungan akan diolah dengan menggunakan tools *rapid miner* sehingga didapatkan grafik yang dapat dilihat pada Gambar 1.



Gambar 1. Hasil *Clustering* rute Perjalanan Dengan *Tools Rapid Miner*

Berdasarkan perhitungan secara manual dan menggunakan *tools rapid miner* seperti pada Gambar 1. didapatkan bahwa *clustering* memiliki hasil pengelompokan yang sama yaitu C0 merupakan *cluster* yang paling diminati terdiri dari rute Tangerang menuju Pantai Anyer, Bandung dan Jogja, sedangkan C1 merupakan rute yang sedang diminati terdiri dari rute Tangerang menuju Bromo, Malang, Kediri dan Blitar, selanjutnya untuk C2 merupakan rute yang kurang diminati terdiri dari rute Tangerang menuju Puncak, Kebun Binatang Ragunan, Banten, Cisauk dan Tangkuban Perahu.

D. Simpulan

Berdasarkan hasil penelitian dapat disimpulkan bahwa dengan metode *clustering* dan algoritma *k-means* dapat menyelesaikan masalah pengelompokan rute perjalanan dimana rute perjalanan dapat dikelompokkan menjadi tiga *cluster* yaitu *cluster* paling diminati (C0) sebesar 20%, *cluster* sedang diminati (C1) sebesar 30% dan *cluster* kurang diminati (C2) sebesar 50%. Dari hasil kesimpulan tersebut manajemen perusahaan menjadi lebih mudah dalam menentukan strategi pemasaran guna meningkatkan jumlah konsumen yang menggunakan jasa perjalanan. Fasilitas serta pelayanan dapat lebih ditingkatkan terutama pada kelompok *cluster* yang kurang diminati konsumen.

E. Referensi

- [1] BPS RI, *Statistik Wisatawan Nusantara 2018*. Jakarta: BPS RI, 2018.
- [2] I. N. Y. Setyawan, K. D. B. A. Anggara, I. P. G. K. Styawan, I. G. W. Pramarta, and G. Indrawan, "Pengelompokan Pelanggan Menggunakan Klaster Dinamik Pada Algoritma K-Means Guna Menentukan Paket Wisata Masa Depan," *Pros. SNRT (Seminar Nas. Ris. Ter. Politek. Negeri Banjarmasin)*, pp. 18–27, 2016.
- [3] N. D. Saksono, Y. A. Sari, and R. K. Dewi, "Rekomendasi Lokasi Wisata Kuliner Menggunakan Metode K-Means Clustering Dan Simple Additive Weighting," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 10, pp. 3835–3842, 2018.
- [4] M. B. Karathiya, R. S. Sakshi, D. S. Sakshi, and D. R. Kathiriya, "Data mining for travels and tourism," *J. Inf. Oper. Manag.*, vol. 3, no. 1, pp. 114–118, 2012.
- [5] P. D. K. P. Fonseka, S. J. Kaluarachchi, D. D. M. Jayaneththi, and T. D. B. S. P. N. Vithana, "Tripper Tourism Management System Using Data Mining," *Int. J. Sci. Res. Publ.*, vol. 7, no. 7, pp. 374–378, 2017.
- [6] N. Iswandhani and M. Muhajir, "K-Means Cluster Analysis of Tourist Destination in Special Region of Yogyakarta using Spatial Approach and Social Network Analysis (a case study : post of @ explorejogja instagram account in 2016)," in *International Conference on Mathematics: Pure, Applied and Computation*, 2018, pp. 1–8.
- [7] R. W. Sari and D. Hartama, "Data Mining: Algoritma K-Means Pada Pengelompokan Wisata Asing ke Indonesia Menurut Provinsi," in *Seminar Nasional Sains & Teknologi Informasi (SENSASI)*, 2018, pp. 322–326.
- [8] J. O. Ong, "Implementasi Algoritma K-Means Clustering Untuk Menentukan Strategi Marketing," *J. Ilm. Tek. Ind.*, vol. 12, no. 1, pp. 10–20, 2013.
- [9] A. Bastian, H. Sujadi, and G. Febrianto, "Penerapan Algoritma K-Means Clustering Analysis Pada Penyakit Menular Manusia (Studi Kasus Kabupaten Majalengka)," *Sist. Inf.*, vol. 1, no. 14, pp. 26–32, 2018.

-
- [10] S. Agrawal, S. Vishwakarma, K, and A. Sharma, K, "Using Data Mining Classifier for Predicting Student ' s Performance in UG Level," *Int. J. Comput. Appl.*, vol. 172, no. 8, pp. 39–44, 2017.
- [11] P. Juwattanasamran, S. Supattranuwong, and S. Sinthupinyo, "Applying Data Mining to Analyze Travel Pattern in Searching Travel Destination Choices," *Int. J. Eng. Sci.*, vol. 2, no. 4, pp. 38–44, 2013.
- [12] V. Mangla, C. Sarda, and T. Nadu, "Improving the efficiency of Apriori Algorithm in Data Mining," *Int. J. Sci. Eng. Technol.*, vol. 3, no. 3, pp. 393–396, 2013.
- [13] J. Han and M. Kamber, *Data Mining Concepts and technique*, Third Edit. United States of America: Elsevier, 2012.
- [14] Z. Arifin, *Penelitian Pendidikan*. Bandung: PT. Remaja Rosdakarya, 2012.