



Hak cipta dan penggunaan kembali:

Lisensi ini mengizinkan setiap orang untuk menggubah, memperbaiki, dan membuat ciptaan turunan bukan untuk kepentingan komersial, selama anda mencantumkan nama penulis dan melisensikan ciptaan turunan dengan syarat yang serupa dengan ciptaan asli.

Copyright and reuse:

This license lets you remix, tweak, and build upon work non-commercially, as long as you credit the origin creator and license it on your new creations under the identical terms.

***CLUSTERING OBJEK PARIWISATA DI PROVINSI
BALI MENGGUNAKAN ALGORITMA K-MEANS DAN
X-MEANS CLUSTERING***

SKRIPSI



Diajukan Guna Memenuhi Persyaratan Memperoleh
Gelar Sarjana Komputer (S.Kom.)

Stephanie Monica

14110310031

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG**

2018

PERNYATAAN

Dengan ini saya menyatakan bahwa skripsi ini adalah karya ilmiah saya sendiri, bukan plagiat dari karya ilmiah yang ditulis oleh orang lain atau lembaga lain, dan semua karya ilmiah orang lain atau lembaga lain yang dirujuk dalam skripsi ini telah disebutkan sumber kutipannya serta dicantumkan di Daftar Pustaka.

Jika di kemudian hari terbukti ditemukan kecurangan / penyimpangan, baik dalam pelaksanaan skripsi maupun dalam penulisan laporan skripsi, saya bersedia menerima konsekuensi dinyatakan TIDAK LULUS untuk mata kuliah skripsi yang telah saya tempuh dan status keserjanaan strata satu yang sudah diterima akan dicabut.

Tangerang, 10 April 2018



Stephanie Monica

HALAMAN PENGESAHAN

Skripsi dengan judul

“*Clustering* Objek Pariwisata di Provinsi Bali Menggunakan Algoritma *K-Means*
dan *X-Means Clustering*”

oleh

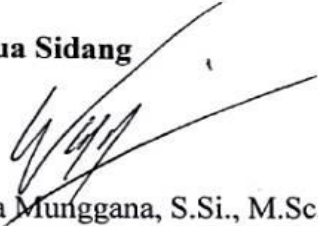
Stephanie Monica

telah diujikan pada hari Selasa, 24 April 2018,


pukul 09.00 s.d. 10.30 dan dinyatakan lulus

dengan susunan penguji sebagai berikut.

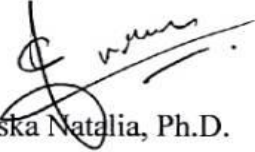
Ketua Sidang


Wira Mungguna, S.Si., M.Sc.

Penguji



Yustinus Eko Soelistio, S.Kom., M.M.

Dosen Pembimbing


Friska Natalia, Ph.D.

Disahkan oleh

Ketua Program Studi Sistem Informasi – UMN

 7/5'18

Ririn Ikana Desanti, S.Kom., M.Kom.

CLUSTERING OBJEK PARIWISATA DI PROVINSI BALI MENGGUNAKAN ALGORITMA *K-MEANS* DAN *X-MEANS CLUSTERING*

ABSTRAK

Oleh: Stephanie Monica

Ada 133 objek pariwisata yang tersebar di provinsi Bali. Untuk itu, dilakukan penelitian mengenai tingkat pengunjung pada objek pariwisata di provinsi Bali. Penelitian dilakukan dengan melakukan *clustering* objek pariwisata di provinsi Bali berdasarkan jumlah pengunjung pada tahun 2012 hingga 2016 menggunakan algoritma *K-Means Clustering* dan *X-Means Clustering*.

Namun ditemukan 18 data objek pariwisata yang memiliki jumlah pengunjung untuk setiap tahun dari tahun 2012 hingga 2016 yaitu 0. Sehingga data dibagi menjadi 2 skenario, skenario 1 menggunakan data objek pariwisata utuh (133 data) dan skenario 2 menggunakan data yang telah dikurangi 18 objek pariwisata yang tidak memiliki jumlah pengunjung untuk tahun 2012 hingga 2016 (115 data).

Hasil *clustering* untuk skenario 1 *K-Means* yaitu 3 *cluster*, skenario 2 *K-Means* yaitu 2 *cluster*, skenario 1 *X-Means* yaitu 4 *cluster*, dan skenario 2 *X-Means* yaitu 4 *cluster*.

Dalam penelitian ini disimpulkan untuk menggunakan algoritma *X-Means Clustering* karena hasil *cluster* lebih merata dan dipilih skenario 1 karena penggunaan data yang lebih lengkap.

Kata kunci: *dashboard*, *K-Means*, visualisasi data, *X-Means*

UMMN

CLUSTERING TOURISM OBJECT IN BALI PROVINCE USING K-MEANS AND X-MEANS CLUSTERING ALGORITHM

ABSTRACT

By: Stephanie Monica

There are 133 tourism objects spread throughout the Bali province. This research conducted to analyze the number of visitors on tourism objects in the Bali province by clustering the tourism objects based on its visitor number in 2012 until 2016 using K-Means Clustering and X-Means Clustering.

However, 18 data of tourist objects founded with total number of visitors of each year from 2012 to 2015 are 0. So the data is divided into 2 scenarios, scenario 1 is using complete data of tourist objects (133 data) and scenario 2 is using data that has reduced by 18 data of tourist objects that have no number of visitors for 2012 to 2016 (115 data).

The clustering result for scenario 1 K-Means is 3 clusters, scenario 2 K-Means is 2 clusters, scenario 1 X-Means is 4 clusters, and scenario 2 X-Means is 4 clusters.

In this research, it is concluded to use X-Means Clustering algorithm because the cluster result is more evenly distributed and scenario 1 is chosen because of the use of more complete data.

Keywords: dashboard, data visualization, K-Means, X-Means

U M N

KATA PENGANTAR

Puji dan syukur kepada Tuhan Yang Esa sehingga penulis dapat menyelesaikan penelitian skripsi yang berjudul “*Clustering* Objek Pariwisata di Provinsi Bali Menggunakan Algoritma *K-Means* dan *X-Means Clustering*” tepat pada waktunya.

Dengan berakhirnya proses dan penulisan penelitian skripsi ini, penulis ingin mengucapkan terima kasih kepada seluruh pihak yang telah mendukung dan membantu penulis untuk menyelesaikan penelitian ini.

Penulis juga ingin mengucapkan terima kasih kepada:

1. Ibu Ririn Ikana Desanti, S.Kom., M.Kom., selaku Ketua Program Studi Sistem Informasi yang telah mendukung penulis dalam pelaksanaan penelitian.
2. Ibu Friska Natalia, Ph.D., selaku dosen pembimbing penelitian skripsi yang telah memberikan bimbingan dan saran selama penulisan laporan skripsi.
3. Seluruh pihak Dinas Pariwisata Provinsi Bali, khususnya Bapak AA. Gede Yuniartha Putra, SH. MH yang telah memberikan izin untuk melakukan penelitian pada Dinas Pariwisata Provinsi Bali, Bapak Ida Bagus Alit Dharmaja, SE yang telah membantu mengarahkan penulis selama penulis berada di kantor Dinas Pariwisata Provinsi Bali, serta

Ibu Dra. Putu Widyahari, M.Si yang telah memberikan data objek pariwisata Bali yang digunakan sebagai bahan penelitian ini.

4. Pihak Pemerintah Provinsi Bali (Dinas Penanaman Modal dan Pelayanan Terpadu Satu Pintu) yang telah memberikan izin bagi penulis untuk melakukan penelitian di provinsi Bali.
5. Kedua orang tua serta keluarga penulis yang telah membesarkan penulis hingga jenjang pendidikan ini. Terima kasih atas didikan, dukungan, dan motivasi yang telah diberikan hingga saat ini.
6. Teman-teman angkatan 2014 yang telah menjadi sahabat sekaligus keluarga dalam menuntun ilmu, serta dukungan dan masukkan saat penulis melakukan penelitian ini.
7. Teman-teman di Mirai Apps yang selalu memberikan dukungan dan semangat saat penulis melaksanakan penelitian.

Penulis meminta maaf apabila terdapat kesalahan dalam pembuatan laporan penelitian skripsi ini. Akhir kata, penulis berharap laporan penelitian skripsi ini dapat bermanfaat bagi para pembaca.

Tangerang, 10 April 2018

Penulis

DAFTAR ISI

PERNYATAAN.....	ii
HALAMAN PERSETUJUAN.....	iii
ABSTRAK.....	iv
<i>ABSTRACT</i>	v
KATA PENGANTAR.....	vi
DAFTAR ISI.....	viii
DAFTAR GAMBAR.....	xi
DAFTAR TABEL.....	xv
DAFTAR RUMUS.....	xvii
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang.....	1
1.2. Rumusan Masalah.....	2
1.3. Batasan Masalah.....	2
1.4. Tujuan dan Manfaat.....	3
1.5. Sistematika Penulisan.....	3
BAB II LANDASAN TEORI.....	5
2.1. <i>Knowledge Discovery in Database (KDD)</i>	5
2.2. <i>Metode Clustering</i>	7
2.3. <i>Algoritma X-Means Clustering</i>	8
2.4. <i>Bayesian Information Criterion (BIC)</i>	8
2.5. <i>Davies-Bouldin Index (DBI)</i>	9
2.6. <i>Algoritma K-Means Clustering</i>	11
2.7. <i>Principal Component Analysis</i>	13

2.8.	Visualisasi Data	13
2.8.1.	<i>Slicer</i>	13
2.8.2.	<i>Clustered Column Chart</i>	14
2.8.3.	<i>Table</i>	14
2.9.	<i>Dashboard</i>	14
2.10.	RapidMiner	15
2.11.	Microsoft Power BI	15
BAB III METODOLOGI PENELITIAN		16
3.1.	Objek Penelitian	16
3.2.	Metode Penelitian.....	16
3.2.1.	Metode Penyelesaian Masalah	18
3.2.1.1.	Metode Penentuan Jumlah <i>Cluster</i> Optimal	18
3.2.1.2.	Metode <i>Clustering</i>	19
3.2.2.	Metode Perancangan Visualisasi Data	20
3.3.	Teknik Pengumpulan Data	23
3.4.	Variabel Penelitian	23
3.3.1.	Variabel Independen (Variabel X).....	23
3.3.2.	Variabel Dependen (Variabel Y)	24
3.5.	Teknik Pengolahan Data	24
3.4.1.	<i>Tools</i> untuk <i>Data Mining</i>	24
3.4.2.	<i>Tools</i> untuk <i>Data Preprocessing</i> dan Visualisasi Data.....	25
3.6.	Penelitian Terdahulu.....	26
BAB IV ANALISIS DAN HASIL PENELITIAN		28
4.1.	<i>Selection</i>	28
4.2.	<i>Preprocessing/Cleaning</i>	37

4.3.	<i>Transformation</i>	55
4.4.	<i>Data Mining</i>	58
4.4.1.	Implementasi Algoritma <i>K-Means Clustering</i>	59
2.4.1.1.	Skenario 1	59
2.4.1.2.	Skenario 2	74
4.4.2.	Implementasi Algoritma <i>X-Means Clustering</i>	89
2.4.2.1.	Skenario 1	89
2.4.2.2.	Skenario 2	98
4.5.	<i>Intepretation/Evaluation</i>	107
4.5.1.	<i>Dashboard</i> Skenario 1 <i>K-Means Clustering</i>	107
4.5.2.	<i>Dashboard</i> Skenario 2 <i>K-Means Clustering</i>	112
4.5.3.	<i>Dashboard</i> Skenario 1 <i>X-Means Clustering</i>	118
4.5.4.	<i>Dashboard</i> Skenario 2 <i>X-Means Clustering</i>	125
4.5.5.	<i>Dashboard</i> Peringkat Objek Pariwisata Terendah	131
4.5.6.	<i>Maintenance</i>	138
4.6.	Analisa dan Diskusi	139
BAB V KESIMPULAN DAN SARAN.....		144
5.1.	Kesimpulan.....	144
5.2.	Saran	146
DAFTAR PUSTAKA		147
LAMPIRAN.....		153

DAFTAR GAMBAR

Gambar 3.1. Tahapan Utama Penelitian	17
Gambar 3.2. <i>Flowchart</i> Tahap dalam <i>Knowledge Discovery in Database</i> (KDD)	21
Gambar 4.1. Grafik Sebaran Jumlah Kunjungan Wisatawan pada Setiap Kabupaten/Kota di Bali Tahun 2012	29
Gambar 4.2. Grafik Sebaran Jumlah Kunjungan Wisatawan pada Setiap Kabupaten/Kota di Bali Tahun 2013	30
Gambar 4.3. Grafik Sebaran Jumlah Kunjungan Wisatawan pada Setiap Kabupaten/Kota di Bali Tahun 2014	32
Gambar 4.4. Grafik Sebaran Jumlah Kunjungan Wisatawan pada Setiap Kabupaten/Kota di Bali Tahun 2015	33
Gambar 4.5. Grafik Sebaran Jumlah Kunjungan Wisatawan pada Setiap Kabupaten/Kota di Bali Tahun 2016	35
Gambar 4.6. Grafik Sebaran Jumlah Kunjungan Wisatawan pada Setiap Kabupaten/Kota di Bali Tahun 2012-2016	36
Gambar 4.7. Tampilan Data 2012 di <i>Query Editor</i>	37
Gambar 4.8. Tampilan Data 2012 Setelah Dihapus Baris dan Kolom yang Tidak Digunakan	38
Gambar 4.9. Tampilan Data 2013 di <i>Query Editor</i>	39
Gambar 4.10. Tampilan Data 2013 Setelah Dihapus Baris dan Kolom yang Tidak Digunakan	40
Gambar 4.11. Tampilan Data 2014 di <i>Query Editor</i>	40
Gambar 4.12. Tampilan Data 2014 Setelah Dihapus Baris dan Kolom yang Tidak Digunakan	41
Gambar 4.13. Tampilan Data 2018 di <i>Query Editor</i>	42

Gambar 4.14. Tampilan Data 2015 Setelah Dihapus Baris dan Kolom yang Tidak Digunakan	43
Gambar 4.15. Tampilan Data 2016 di <i>Query Editor</i>	43
Gambar 4.16. Tampilan Data 2016 Setelah Dihapus Baris dan Kolom yang Tidak Digunakan	44
Gambar 4.17. Tampilan Data 2012 Setelah Nama Objek Disamakan.....	45
Gambar 4.18. Tampilan Data 2013 Setelah Nama Objek Disamakan.....	46
Gambar 4.19. Tampilan Data 2014 Setelah Nama Objek Disamakan.....	46
Gambar 4.20. Tampilan Data 2015 Setelah Nama Objek Disamakan.....	47
Gambar 4.21. Tampilan Data 2016 Setelah Nama Objek Disamakan.....	48
Gambar 4.22. Tabel Total Pengunjung	49
Gambar 4.23. Tabel Daftar Objek.....	50
Gambar 4.24. Tabel Total Pengunjung 1	52
Gambar 4.25. Tabel Daftar Objek 1	52
Gambar 4.26. Tabel Total Pengunjung 2	53
Gambar 4.27. Tabel Daftar Objek 2.....	54
Gambar 4.28. <i>Flowchart</i> proses <i>transformation</i>	55
Gambar 4.29. Tabel Total Pengunjung 1 dalam Format <i>Excel</i>	56
Gambar 4.30. Tabel Total Pengunjung 2 dalam Format <i>Excel</i>	56
Gambar 4.31. Tabel Daftar Objek 1 dalam Format <i>Excel</i>	57
Gambar 4.32. Tabel Daftar Objek 1 dalam Format <i>Excel</i>	58
Gambar 4.33. Proses Skenario 1 <i>K-Means Clustering</i>	59
Gambar 4.34. Subproses Skenario 1 <i>K-Means Clustering</i>	59
Gambar 4.35. Parameter Operator <i>Log</i> dalam Subproses Skenario 1 <i>K-Means Clustering</i>	66

Gambar 4.36. Parameter Operator <i>Loop Parameters</i> dalam Proses Skenario 1 <i>K-Means Clustering</i>	67
Gambar 4.37. Proses dalam Skenario 2 <i>K-Means Clustering</i>	74
Gambar 4.38. Subproses dalam Skenario 2 <i>K-Means Clustering</i>	75
Gambar 4.39. Parameter Operator <i>Log</i> dalam Skenario 2 <i>K-Means Clustering</i> ...	82
Gambar 4.40. Parameter Operator <i>Loop Parameters</i> dalam Skenario 2 <i>K-Means Clustering</i>	83
Gambar 4.41. Proses dalam Skenario 1 <i>X-Means Clustering</i>	90
Gambar 4.42. Proses dalam Skenario 2 <i>X-Means Clustering</i>	99
Gambar 4.43. <i>Dashboard K-Means</i> Skenario 1	108
Gambar 4.44. Visualisasi <i>Slicer Cluster Dashboard K-Means</i> Skenario 1	109
Gambar 4.45. Visualisasi <i>Slicer Tipe Objek Dashboard K-Means</i> Skenario 1 ..	109
Gambar 4.46. Visualisasi <i>Slicer Kabupaten/Kota Dashboard K-Means</i> Skenario 1	110
Gambar 4.47. Visualisasi Rata-Rata Pengunjung Setiap <i>Cluster K-Means</i> Skenario 1	111
Gambar 4.48. Visualisasi Jumlah Pengunjung Per Tahun Pada Setiap Objek Pariwisata <i>K-Means</i> Skenario 1	112
Gambar 4.49. <i>Dashboard K-Means</i> Skenario 2	114
Gambar 4.50. Visualisasi <i>Slicer Cluster Dashboard K-Means</i> Skenario 2	115
Gambar 4.51. Visualisasi <i>Slicer Tipe Objek Dashboard K-Means</i> Skenario 2 ..	115
Gambar 4.52. Visualisasi <i>Slicer Kabupaten/Kota Dashboard K-Means</i> Skenario 2	116
Gambar 4.53. Visualisasi Rata-Rata Pengunjung Setiap <i>Cluster K-Means</i> Skenario 2	117
Gambar 4.54. Visualisasi Jumlah Pengunjung Per Tahun Pada Setiap Objek Pariwisata <i>K-Means</i> Skenario 2	118

Gambar 4.55. <i>Dashboard X-Means</i> Skenario 1	120
Gambar 4.56. Visualisasi <i>Slicer Cluster Dashboard X-Means</i> Skenario 1	121
Gambar 4.57. Visualisasi <i>Slicer Tipe Objek Dashboard X-Means</i> Skenario 1 ..	121
Gambar 4.58. Visualisasi <i>Slicer Kabupaten/Kota Dashboard X-Means</i> Skenario 1	122
Gambar 4.59. Visualisasi Pengunjung Setiap <i>Cluster X-Means</i> Skenario 1	123
Gambar 4.60. Visualisasi Jumlah Pengunjung Per Tahun Pada Setiap Objek Pariwisata <i>X-Means</i> Skenario 1	124
Gambar 4.61. <i>Dashboard X-Means</i> Skenario 2	126
Gambar 4.62. Visualisasi <i>Slicer Cluster Dashboard X-Means</i> Skenario 2	127
Gambar 4.63. Visualisasi <i>Slicer Tipe Objek Dashboard X-Means</i> Skenario 2 ..	127
Gambar 4.64. Visualisasi <i>Slicer Kabupaten/Kota Dashboard X-Means</i> Skenario 2	128
Gambar 4.65. Visualisasi Rata-Rata Pengunjung Setiap <i>Cluster X-Means</i> Skenario 2	129
Gambar 4.66. Visualisasi Jumlah Pengunjung Per Tahun Pada Setiap Objek Pariwisata <i>X-Means</i> Skenario 2	130
.....	132
Gambar 4.67. <i>Dashboard</i> Peringkat Objek Pariwisata Terendah	132
Gambar 4.68. Visualisasi <i>Slicer Tahun Dashboard</i> Peringkat Objek Pariwisata Terendah.....	133
Gambar 4.69. Visualisasi <i>Table</i> Objek Pariwisata <i>Dashboard</i> Peringkat Objek Pariwisata Terendah	134

DAFTAR TABEL

Tabel 3.1. Perbandingan antara <i>Bayesian Information Criterion</i> (BIC) dan <i>Davies-Bouldin Index</i> (DBI).....	18
Tabel 3.2. Perbandingan antara Algoritma <i>X-Means Clustering</i> dan Algoritma <i>K-Means Clustering</i>	19
Tabel 3.3. Perbandingan antara KDD dengan CRISP-DM.....	20
Tabel 3.4. Perbandingan antara RapidMiner dengan MATLAB	24
Tabel 3.5. Perbandingan antara <i>Microsoft Power BI</i> dengan <i>Tableau</i>	25
Tabel 3.6. Penelitian Terdahulu	26
Tabel 4.1. <i>Attribute</i> dan <i>Role</i> Data Total Pengunjung 1 pada Proses Skenario 1 <i>K-Means Clustering</i>	60
Tabel 4.2. Parameter Operator PCA pada Proses Skenario 1 <i>K-Means Clustering</i>	61
Tabel 4.3. Parameter Operator <i>K-Means</i> dalam Subproses Skenario 1 <i>K-Means Clustering</i>	62
Tabel 4.4. Parameter Operator <i>Cluster Distance Performance</i> dalam Subproses Skenario 1 <i>K-Means Clustering</i>	65
Tabel 4.5. Tabel <i>Log</i> Skenario 1 <i>K-Means Clustering</i>	68
Tabel 4.6. Parameter Operator <i>K-Means</i> pada Proses Skenario 1 <i>K-Means Clustering</i>	69
Tabel 4.7. <i>Attribute</i> dan <i>Role</i> Data Total Pengunjung 1 pada Proses Skenario 2 <i>K-Means Clustering</i>	75
Tabel 4.8. Parameter Operator PCA pada Proses Skenario 2 <i>K-Means Clustering</i>	76
Tabel 4.9. Parameter Operator <i>K-Means</i> pada Subproses Skenario 2 <i>K-Means Clustering</i>	78

Tabel 4.10. Parameter Operator <i>Cluster Distance Performance</i> pada Subproses Skenario 2 <i>K-Means Clustering</i>	81
Tabel 4.11. Hasil Operator <i>Log</i> pada Subproses Skenario 2 <i>K-Means Clustering</i>	83
Tabel 4.12. Parameter Operator <i>K-Means</i> pada Proses Skenario 2 <i>K-Means Clustering</i>	84
Tabel 4.13. <i>Attribute</i> dan <i>Role</i> Data Total Pengunjung 1 pada Proses Skenario 1 <i>K-Means Clustering</i>	90
Tabel 4.14. Parameter Operator PCA pada Proses Skenario 1 <i>K-Means Clustering</i>	91
Tabel 4.15. Parameter Operator <i>X-Means</i> pada Proses Skenario 1 <i>X-Means Clustering</i>	92
Tabel 4.16. <i>Attribute</i> dan <i>Role</i> Data Total Pengunjung 1 pada Proses Skenario 1 <i>K-Means Clustering</i>	99
Tabel 4.17. Parameter Operator PCA pada Proses Skenario 1 <i>K-Means Clustering</i>	100
Tabel 4.18. Parameter Operator <i>X-Means</i> pada Proses Skenario 2 <i>X-Means Clustering</i>	101
Tabel 4.19. Perbedaan Sebelum dan Sesudah Dilakukan Penelitian	143



DAFTAR RUMUS

Rumus 2.1. <i>Bayesian Information Criterion (BIC)</i>	9
Rumus 2.2. <i>Sum of Square within Cluster (SSW)</i>	9
Rumus 2.3. <i>Sum of Square between Cluster (SSB)</i>	10
Rumus 2.4. Pengukuran Nilai Rasio	10
Rumus 2.5. <i>Davies-Bouldin Index</i>	11
Rumus 2.5. <i>Euclidean</i>	12

UMMN