# Principal Component Analysis untuk Dimensi Reduksi Data Clustering Sebagai Pemetaan Persentase Sertifikasi Guru di Indonesia

ISSN (Printed) : 2579-7271 ISSN (Online) : 2579-5406

# Gusni Rahayu<sup>1</sup>, Mustakim<sup>2</sup>

<sup>1,2</sup>Laboratorium Data Mining Program Studi Sistem Informasi, Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau
JI.HR. Soebrantas No. 155 Simpang Baru, Tampan, Pekanbaru, Riau – Indonesia 28293
Email: gusni.rahayu@students.uin-suska.ac.id, mustakim@uin-suska.ac.id

#### Abstrak

Penelitian ini bertujuan untuk mengetahui kota-kota mana saja yang akan digunakan untuk klasterisasi wilayah di Indonesia berdasarkan tingkat pengajar yang sudah profesional yaitu melalui sertifikasi. Data yang digunakan terdiri dari 3 jenjang pendidikan yaitu SD,SMP dan SMA yang ada diseluruh kota di Indonesia sebanyak 98 kota. Pendidikan tidaklah diarahkan hanya dalam mencetak tenaga kerja untuk industri melainkan juga tenaga kerja yang mengoptimalkan kemampuan berpikir dalam menjalankan pekerjaanya. Proses pemetaan dapat dilakukan dengan menggunakan salah satu teknik clustering yaitu K-Means. K-Means merupakan algoritma paling populer yang digunakan karena memiliki kelebihan yaitu algoritma yang sederhana dan mudah diimplementasikan. Principal Component Analysis (PCA) merupakan pendekatan fitur selection untuk pengurangan dimensi tanpa pengawasan teknik. Penelitian ini menghasilkan sebuah metode yang lebih efektif dengan menggunakan PCA yang dimodifikasi dengan k-means.Dari pembagian cluster sebanyak K=3 didapatkan hasil cluster 0 = 13, cluster 1 = 44, cluster 2 = 41. Metode pengukuran Davies-Bouldin (DB) Index digunakan untuk mengukur validitas cluster yang kemudian menghasilkan algoritma K-Means memiliki nilai yang lebih besar dari pada PCA-K-Means.

Kata Kunci: Dimension reduction, K-means, Cluster, Principal Component Analysis, Sertifikasi guru

### Abstract

This study aims to determine which cities are going to be used for clustering regions in Indonesia based on the teachers who are professionals, namely through certification. The data used consists of three levels of education is elementary, junior and senior high schools through the towns in Indonesia as many as 98 cities. Education is not directed only in the printing industry labor but also to optimize the ability to think in carrying out his job. The mapping process can be done by using one of the techniques that K-Means clustering. K-Means is the most popular algorithms used because it has the advantages of simple algorithm and easy to implement. Principal Component Analysis (PCA) is an approach to the reduction of the dimensions of feature selection techniques without supervision. The study produced a more effective method of using PCA modified with k-means.Dari cluster division as K=3 is obtained cluster 0=13, 1=44 clusters, cluster 2=41. The measurement method of Davies-Bouldin (DB) Index is used to measure the validity of the cluster which then produces the K-Means algorithm has a greater value than the PCA-K-Means.

Key words: Dimension reduction, K-means, Cluster, Principal Component Analysis, Teacher certification

## 1. Pendahuluan

Peningkatan kualitas sumber daya manusia merupakan salah satu penekanan dari tujuan pendidikan yang tertuang dalam undang-undang 1945 dalam alenia ke-4. Perkembangan kebutuhan masyarakat atas sumber daya manusia yang berkualitas, perlahan namun pasti semakin meningkat dari tahun ke tahun. Hal ini sejalan dengan perkembangan tuntutan dunia kerja yang tidak hanya membutuhkan sumber daya manusia yang berorientasi untuk kebutuhan dunia industri. Sumber daya manusia yang dibutuhkan saat ini adalah sumber daya manusia yang memiliki kompetensi unggulan terutama dalam hal kemampuan berpikir. Untuk menjadi manusia-manusia yang berkompeten maka sumber daya manusia tersebut haruslah memperoleh pendidikan terlebih dahulu [15].

Program sertifikasi pendidik dilaksanakan dalam rangka memenuhi amanat Undang-Undang Republik Indonesia Nomor 20 Tahun 2003 tentang Sistem Pendidikan Nasional, Undang-Undang Republik Indonesia Nomor 14 Tahun 2005 tentang Guru dan Dosen, dan Peraturan Pemerintah Nomor 19 Tahun 2005 tentang Standar Nasional Pendidikan. Sertifikasi

guru merupakan kebijakan yang sangat strategis, karena langkah dan tujuan melakukan sertifikasi guru untuk meningkat kualitas guru, memiliki kompetensi, mengangkat harkat dan wibawa guru sehingga guru lebih dihargai dan untuk meningkatkan kualitas pendidikan di Indonesia [1].Oleh karena itu, perlu dilakukan pemetaan wilayah-wilayah yang ada di Indonesia menggunakan salah satu teknik *clustering* yaitu *K-Means* untuk dijadikan evaluasi terhadap kebijakan pemerintah dalam menentukan seberapa banyak guru yang sudah memenuhi program pendidikan di Indonesia yang dapat meningkatkan mutu pendidikan dan daerah mana saja yang memiliki tenaga pengajar yang sudah berkualitas mulai dari SD,SMP dan SMA. Pengelompokan ini bertujuan untuk menemukan pola kedekatan terkait lebaran data kota di Indonesia yang memiliki ciri yang sama terkait data sertifikasi. Hasil-hasil dari pemodelan yang dihasilkan mampu menemukan struktur kesamaan antar data dari masing-masing kota terhadap lebaran data yang dilakukan pada pengolahan sebuah *Data Mining*.

ISSN (Printed) : 2579-7271 ISSN (Online) : 2579-5406

Data Mining merupakan proses untuk menyelesaikan permasalahan diatas. Algoritma K-Means merupakan teknik *clustering* pada data mining yang memiliki beberapa kelebihan, diantaranya adalah proses *cluster* bisa dilakukan dengan cepat karena memiliki beban komputasi relatif lebih ringan [7].

K-Means memiliki beberapa kelemahan salah satunya yaitu penentuan pusat awal cluster. Hasil cluster yang terbentuk dari metode K-Means ini tergantung pada inisiasi nilai pusat awal cluster yang diberikan. Hal ini menyebabkan hasil cluster nya berupa solusi yang sifatnya local optimal. Algoritma klasifikasi dan clustering menjadi bermasalah pada data dengan dimensi tinggi berupa menurunnya akurasi klasifikasi dan kualitas cluster yang jelek dan juga berpengaruh pada waktu komputasi yang lama. Permasalahan ini dapat dapat diatasi dengan melakukan reduksi dimensi [5].

Ada kelebihan yang didapat dengan melakukan reduksi dimensi. *Principal Component Analysis* (PCA) adalah salah satu fitur reduksi variabel yang banyak digunakan dalam statistika *multivariate*. Tujuan dari PCA adalah dapat mereduksi dimensi data yang tinggi menjadi dimensi data yang lebih rendah dengan resiko kehilangan informasi yang sangat kecil [5]. Oleh karena itu, untuk meningkatkan efisiensi,digunakan sebuah metode untuk menerapkan PCA pada set data asli, sehingga variabel berkorelasi ada di dataset asli akan diubah ke variabel mungkin berkorelasi. Sebelum menerapkan PCA dataset perlu dinormalisasi, sehingga setiap atribut dengan domain yang lebih besar tidak akan mendominasi atribut dengan domain yang lebih kecil. Sehingga mengurangi set data yang diperoleh dari penerapan PCA akan diterapkan pada algoritma clustering [8].

Penelitian ini mengusulkan metode PCA untuk mereduksi dimensi untuk optimasi *cluster* pada algoritma *K-Means* untuk pemetaan wilayah guru penerima sertifikasi.

#### 2. Tinjauan Pustaka

## 2.1 Data Mining

Data mining adalah sebuah proses untuk menguraikan penemuan pengetahuan didalam sebuah database. Kemudian Data mining merupakan proses yang menggunakan teknik statistic, matematika, kecerdasan buatan, dan mechine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat bagi pengetahuan yang terkait dari berbagai database dengan kapasitas besar [7]. Data mining dapat dilakukan pada berbagai jenis database dan repositori informasi, tapi jenis pola dapat ditemukan dengan ditentukan oleh berbagai fungsi data mining seperti deskripsi kelas, asosiasi, korelasi analisis, klasifikasi, prediksi, analisis cluster dll [10].

### 2.2 Analisis Clustering

Teknik *clustering* adalah pengelompokan data yang didasarkan hanya pada informasi yang ditemukan dalam data yang menggambarkan objek tersebut dan hubungan diantaranya. Tujuannya adalah agar objek-objek yang bergabung dalam sebuah kelompok merupakan objek-objek yang mirip atau berhubungan satu sama lain dan berbeda atau tidak berhubungan dengan objek kelompok lain. Lebih besar kemiripanya (homogenitas) dalam kelompok dan lebih besar perbedaannya diantara kelompok yang lain [6]. *Clustering* melakukan pengelompokan data yang di dasarkan pada kesamaan antar objek, oleh karena itu klasterisasi digolongkan sebagai metode *unsupervised learning* [12].

### 2.3 Algoritma K-Means

K-Means *Clustering* adalah salah suatu metode *cluster analysis* pada Data Mining yang termasuk non hirarki yang berusaha untuk mempartisi objek yang ada kedalam satu atau lebih *cluster* atau kelompok objek berdasarkan karakteristiknya, sehingga objek yang mempunyai karakteristik yang sama dikelompokan dalam satu *cluster* yang sama dan objek yang mempunyai karakteristik yang berbeda dikelompokan kedalam *cluster* yang berbeda [3].

ISSN (Printed) : 2579-7271 ISSN (Online) : 2579-5406

## 2.4 Principal Component Analysis (PCA)

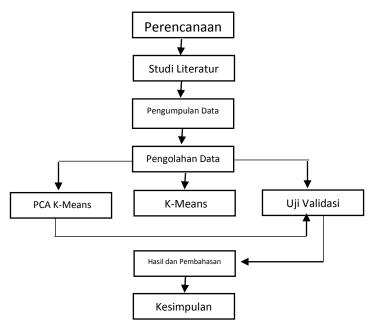
Principal Component Analysis merupakan teknik yang biasa digunakan untuk menyederhanakan suatu data, dengan cara mentransformasi data secara linier sehingga terbentuk sistem koordinat baru dengan varians maksimum. Analisis PCA dapat digunakan untuk mereduksi dimensi suatu data tanpa mengurangi karakteristik data tersebut secara signifikan [11].

#### 2.5 Sertifikasi guru

Sertifikasi adalah proses pemberian sertifikat pendidik untuk Guru dan Dosen atau bukti formal sebagai pengakuan yang diberikan kepada Guru dan Dosen sebagai tenaga profesional [1]. Berdasarkan pengertian tersebut, sertifikasi guru dapat diartikan sebagai suatu proses pemberian pengakuan bahwa seseorang telah memiliki kompetensi untuk melaksanakan pelayanan pendidikan pada satuan pendidikan tertentu, setelah lulus uji kompetensi yang diselenggarakan oleh lembaga sertifikasi.

### 3. Metode Penelitian

Metodologi yang diterapkan dalam melakukan penelitian ditunjukkan pada diagram berikut :



Gambar 1. Metodologi Penelitian

#### 4. Analisa dan Pembahasan

#### 4.1 Pengumpulan Data

Pengumpulan data merupakan tahapan untuk mendapatkan data yang digunakan dalam melakukan penelitian setelah melakukan perencanaan.Pengumpulan data didapatkan dari Pusat Data dan Statistik Pendidikan dan Kebudayaan berjumlah 549 data, yang terdiri dari guru SD,SMP, SMA, PLB, dan SMK. Tidak semua atribut dipakai, tetapi hanya SD, SMP, dan

SMA untuk data persentase guru penerima sertifikasi diseluruh kota yang ada di Indonesia sebanyak 98 data disajikan sebagai berikut:

ISSN (Printed): 2579-7271

ISSN (Online): 2579-5406

Tabel 1. Data persentase sertifikasi guru

No	Nama kota	SD	SMP	SMA
1	Kota Ambon	45,61	54,72	34,43
2	Kota Balikpapan	41,76	52,38	30,39
3	Kota Banda Aceh	61,97	66,36	65,91
4	Kota Bandar Lampung	54,35	59,22	54,94
5	Kota Bandung	52,98	56,85	48,3
6	Kota Banjar	65,43	57,26	50,53
7	Kota Banjarbaru	41,83	54,62	55,71
8	Kota Banjarmasin	41,74	62,31	53,71
9	Kota Batam	23,8	28,09	33
10	Kota Batu	57,6	53,77	46,08
98	Kota Yogyakarta	55,87	64,8	46,64

Langkah selanjutnya adalah melakukan normalisasi data dengan menggunakan min-max *normalization*. Hasil normalisasi ditunjuk pada tabel 2 berikut:

Tabel 2. Normalisasi data

No	Nama kota	SD	SMP	SMA
1	Kota Ambon	0,407	0,556	0,336
2	Kota Balikpapan	0,336	0,508	0,262
3	Kota Banda Aceh	0,713	0,796	0,907
4	Kota Bandar Lampung	0,571	0,649	0,708
5	Kota Bandung	0,545	0,6	0,587
6	Kota Banjar	0,778	0,608	0,628
7	Kota Banjarbaru	0,337	0,554	0,722
8	Kota Banjarmasin	0,335	0,712	0,686
9	Kota Batam	0	0,007	0,31
10	Kota Batu	0,631	0,536	0,547
98	Kota Yogyakarta	0,599	0,764	0,557

## 4.2 Uji kelayakan KMO

Mengenai layak atau tidaknya analisis faktor, maka data kriteria perlu dilakukan uji Kaiser mayer Olkin (KMO) dan Barlett Test. Jika nilai KMO kurang dari 0,5 maka analisis faktor tidak layak dilakukan [6]. Hasil pengujian data menggunakan SPSS 16 dengan hasil pada tabel berikut:

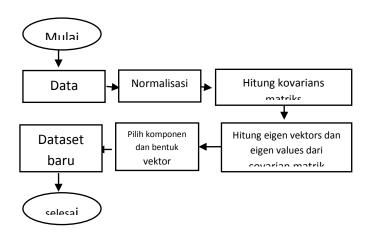
Tabel 3. Hasil Uji KMO

KMO and Bartlett's Test			
Kaiser-Meyer-Olkin Measure	,655		
Bartlett's Test of Sphericity	Approx. Chi-Square	86,440	
	Df	3	
	Sig.	,000	

Berdasarkan tabel diatas, data dapat digunakan untuk proses selanjutnya. Hal ini berdasarkan test KMO dan Bartlett's Test dengan nilai 0,655 dimana Nilai ini sudah mencukupi standart digunakan dalam proses PCA, yaitu dengan nilai > 0,500.

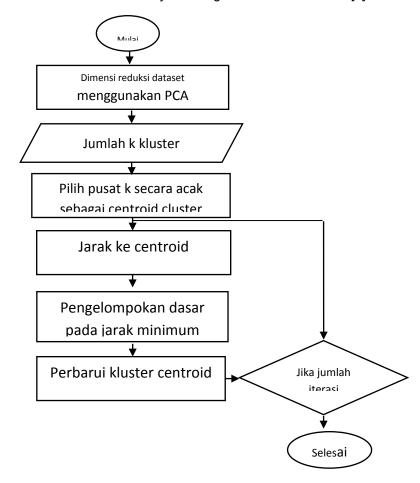
## 4.3 Pengolahan Data

Penelitian ini mengkombinasikan algoritma *K-Means* tradisional dengan metode PCA. Sebelum proses *clustering* dilakukan, reduksi dimensi *dataset* dilakukan menggunakan PCA.Metode PCA dilakukan sesuai dengan tahapan-tahapan yang digambarkan pada Gambar 2 [5].



Gambar 2. Algoritma metode PCA

Sebagai perbandingan, dalam pengujian dilakukan pengukuran terhadap kualitas cluster antara algoritma *K-Means* tradisional dengan algoritma *K-Means* yang telah dikombinasikan dengan metode PCA. Gambar 3 menunjukkan algoritma PCA-*K-Means* [5].



ISSN (Printed): 2579-7271

ISSN (Online): 2579-5406

Gambar 3. Kombinasi PCA-K-Means

Untuk mengevaluasi metode yang diusulkan, penelitian ini akan membandingkan antara metode *K-Means* tradisional (*K-Means*) dengan metode *K-Means* yang diawali dengan proses reduksi dimensi *dataset* menggunakan PCA K-*Means*.

### 4.4 Reduksi dimensi dengan PCA

Pada tahap ini dilakukan reduksi dimensi *dataset* menggunakan PCA. Penelitian ini mereduksi dimensi *dataset* menjadi data 2 dimensi. Hasil reduksi dimensi *dataset* ditunjukkan pada tabel berikut:

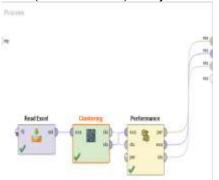
Tabel 4. Reduksi dimensi *dataset* menjadi 2 dimensi menggunakan PCA

No	PC1	PC2
1	-204	-120
2	-313	-128
3	0.442	0.101
4	0.160	0.068
5	0.042	0.007
6	0.179	-472
7	0.001	0.210
8	0.079	0.102
9	-758	0.281
10	0.016	-23
98	0.153	-118

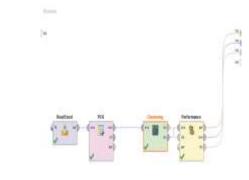
Setelah dilakukan reduksi dimensi *dataset* menggunakan PCA, tahapan berikutnya adalah tahap *clustering*.

## 4.5 Clustering menggunakan Algoritma K-Means

Implementasi algoritma *K-Means* dalam penelitian ini menggunakan alat bantu *Rapid-Miner*. Berturut-turut, pemodelan *K-Means* dan pemodelan *K-Means* yang dikombinasikan dengan PCA (PCA-*K-Means*) ditunjukkan dalam Gambar berikut:



Gambar 4.Pemodelan k-means



ISSN (Printed): 2579-7271

ISSN (Online) : 2579-5406

Gambar 5. Pemodelan PCA-K-means

Dilakukan percobaan terhadap model yang dibangun,yang pertama yaitu *clustering* terhadap *dataset* K-*Means* tradisional dan kemudian terhadap data reduksi dimensi (PCA-K-Means) yang hasil segmentasi yang terbentuk akan dievaluasi menggunakan *Davies-Bouldin* (DB) Index.

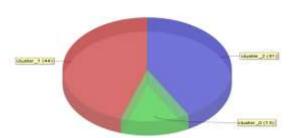
## 4.6 Uji validasi

Uji validasi yang dilakukan pada penelitian ini menggunakan Uji validasi yang dilakukan pada penelitian ini menggunakan *Davis- Bouldin Index* (DBI) dengan para meter semakin kecil nilai DBI maka semakin baik [6]. Uji validasi *cluster* diperoleh hasil sebagai berikut:

Tabel 5. Nilai DB index percobaan

raber of reliai BB index percebaan			
Model clustering	Hasil Uji Validasi		
K-means	-1.158		
PCA-K-Means	-1.035		

Pca- kmeans merupakan algoritma dengan nilai terkecil sehingga dapat disimpulkan bahwa algoritma tersebut lebih baik dibandingkan K-Means. Dari hasil clustering PCA-K-Means diperoleh hasil pengelompokan sebagai berikut:



ISSN (Printed): 2579-7271

ISSN (Online): 2579-5406

Gambar 4. Diagram PCA-K-Means

## 4.7 Analisa Hasil

Dari hasil percobaan yang dilakukan, algoritma *K-Means* dengan PCA.Jumlah *cluster* yang terbentuk adalah 3 dengan segmentasi sebagai berikut :

Cluster 0 : 13 items Cluster 1 : 44 items Cluster 2 : 41 items Total number of item: 98

Cluster 1 merupakan kumpulan wilayah yang memiliki nilai baik untuk semua atribut yang bisa dikelompokan sebagai wilayahh yang banyak guru penerima sertifikasi. Cluster 2 merupakan kelompok wilayah yang memiliki cukup guru sebagai penerima sertifikasi ,dan cluster 0 merupakan kelompok wilayah yang sedikit memiliki guru yang sudah menerima sertifikasi. Hasil cluster wilayah dengan PCA-K-Means adalah sebagai berikut:

Tabel 6. Hasil Cluster 0, 1 dan 2

Cluster	Jumlah item			
		Kota Balikpapan	Kota Sorong	Kota Tangerang Selatan
		Kota Batam	Kota Subulussalam	Kota Ternate
Cluster 0	13	Kota Bekasi	Kota Sungai Penuh	Kota Tual
		Kota Denpasar	Kota Pekalongan	Kota Singkawang
		Kota Jakarta Utara		
		Kota Banda Aceh	Kota Jambi	Kota Pasuruan
		Kota Bandar Lampung	Kota Kediri	Kota Pematang siantar
		Kota Banjar	Kota Kendari	Kota Probolinggo
		Kota Banjarmasin	Kota Kotamobagu	Kota Sawah Lunto
		Kota Baubau	Kota Madiun	Kota Semarang
		Kota Binjai	Kota Magelang	Kota Solok
		Kota Bitung	Kota Manado	Kota Sukabumi
Cluster 1	44	Kota Blitar	Kota Metro	Kota Surakarta
		Kota Bogor	Kota Mojokerto	Kota Tanjung Balai
		Kota Bukittinggi	Kota Padang Panjang	Kota Tasikmalaya
		Kota Cimahi	Kota Padang Sidimpuan	Kota Tebing Tinggi
		Kota Cirebon	Kota Palangka Raya	Kota Tegal
		Kota Dumai	Kota Palu	Kota Tomohon
		Kota Gorontalo	Kota Pare-Pare	Kota Yogyakarta
		Kota Jakarta Timur	Kota Pariaman	
		Kota Ambon	Kota Depok	Kota Pekanbaru
		Kota Bandung	Kota Gunungsitoli	Kota Pontianak
		Kota Banjarbaru	Kota Jakarta Barat	Kota Prabumulih
		Kota Batu	Kota Jakarta Pusat	Kota Sabang
		Kota Bengkulu	Kota Jakarta Selatan	Kota Salatiga
		Kota Bima	Kota Jayapura	Kota Samarinda
Cluster 2	41	Kota Medan	Kota Kupang	Kota Serang
		Kota Padang	Kota Langsa	Kota Sibolga
		Kota Pagar Alam	Kota Lhokseumawe	Kota Surabaya
		Kota Palembang	Kota Lubuk Linggau	Kota Tangerang
		Kota Palopo	Kota Makassar	Kota Tanjungpinang
		Kota Pangkalpinang	Kota Malang	Kota Tarakan
		Kota Bontang	Kota Mataram	Kota Tidore Kepulauan
		Kota Cilegon	Kota Payakumbuh	

## 5. Kesimpulan

Metode PCA dapat digunakan untuk mereduksi dimensi *dataset* tanpa harus kehilangan banyak informasi. Dari penelitian yang telah dilakukan, reduksi dimensi *dataset* menggunakan metode PCA terbukti dapat meningkatkan kualitas *cluster* yang dihasilkan dengan algoritma *K-Means*. Pada percobaan diatas peningkatan kualitas permodelan dapat dilihat dari peningkatan akurasi yang signifikan. Dan di dapatkan hasil cluster 0 sebanyak 13 kota, cluster 1 sebanyak 44 kota dan cluster 2 sebanyak 41 kota. Hasil pengukuran validitas *cluster* dengan *Davies-Bouldin Index* menunjukkan PCA-*K-Means* menghasilkan nilai DB *Index* lebih kecil dibandingkan dengan *K-Means*.

ISSN (Printed) : 2579-7271 ISSN (Online) : 2579-5406

#### Ucapan Terimakasih

Terimakasih kepada Fakultas Sains dan Teknologi dan Program Studi Sistem Informasi Universitas Islam Negri Sultan Syarif Kasim Riau atas segala dukungan fasilitas yang telah diberikan dalam menunjang pembuatan *paper* ini. Selanjutnya Komunitas *Puzzle Research Data Technology* (PREDATECH) atas segala motivasi dan bantuannya sehingga *paper* ini dapat terselesaikan dengan baik.

#### **Daftar Pustaka**

- [1] Budiharto. (2013). Sertifikasi guru sebagai upaya meningkatkan profesionalisme guru dalam rangka meningkatkan kualitas pendidikan. 115-128.
- [2] Caesarendra, W. (2011). panduan belajar mandiri Matlab. Jakarta: PT Elex Media Komputindo.
- [3] Ediyanto, M. N. (2013). Pengklasifikasian Karakteristik dengan Metode K-means Cluster Analysis. Buletin Ilmiah Mat, Stat, dan terapannya (Bimaster), 133-136.
- [4] Fahmida Afrin, M.-A. T. (2015). Comparative Performance Of Using PCA With K-Means And Fuzzy C Means Clustering For Costomer Segmentation. *International Journal Of Scientific & Technology Research*, Volume 4.
- [5] Izzuddin, A. (2015). Optimasi Cluster pada Algoritma K-Means dengan Reduksi Dimensi Dataset menggunakam principal component analysis untuk pemetaan kinerja dosen. vol 5 No.
- [6] Jariadi, A. (April 2016). pengelompokan kecamatan penghasil kelapa sawit di provinsi riau menggunakan principal component analysis k-means. *Tugas Akhir Sistem Informasi*.
- [7] Mustakim. (2012). Pemetaan digital dan pengelompokan lahan hiaju di wilayah provinsi riau berdasarkan knowledge discovery in databases (KDD) dengan teknik k-means mining. Seminar Nasional teknologi informasi,komunikasi dan industri (SNTKI)4.
- [8] Rajashree Dash, D. M. (2010). hybridized K-means clustering approach for high dimensional data. *International Journal of Engineering, Science and Technology*, vol 2,No.2 pp.59-66.
- [9] S.Uyanto, S. (2009). *Pedoman sata dengan SPSS*. Yogyakarta: Graha Ilmu.
- [10] Saravanan, T. d. (2011). An efficient method to improve the clustering performance for high dimensional data by Principal Component analysis and modified K-means. *International Jornal of Database Management System (IJDMS)*, vol.3,No 1.
- [11] Septa Firmansyah Putra, R. P. (2016). Feature Selection pada Dataset Faktor Kesiapan Bencana pada Provinsi di Indonesia Menggunakan metode PCA(Principal Componen Analysis). *Jurnal Teknik ITS*, Vol.5,No.2.
- [12] Setiani, N. (2014). Klasterisasi Judul Buku dengan menggunakan metode K-Means. Seminar Nasional Aplikasi Teknologi informasi (SNATI).
- [13] Stanislaus S.Uyanto, P. (2009). Pedoman Analisis Data dengan SPSS. Yogyakarta: Graha ilmu.
- [14] Sunjoyo, R. s. (2013). Aplikasi SPSS untuk smart Riset. Bandung: Alfabeta.
- [15] Tono. (2014). Hubungan sertifikasi guru dengan kinerja guru sekolah dasar negeri Kecamatan Sebuku Kecamatan Nunukan. e*Jurnal Administrasi negara*, *2*(4).