

Optimasi *Cluster* pada Algoritma *K-Means* dengan Reduksi Dimensi *Dataset* Menggunakan *Principal Component Analysis* untuk Pemetaan Kinerja Dosen

Ahmad Izzuddin

Program Studi Teknik Elektro, Fakultas Teknik, Universitas Panca Marga
Jl. Yos Sudarso 107 Pabean Dringu Probolinggo 67271
Email : oedienpowerful@yahoo.com

Terima Naskah : 01 Mei 2015

Terima Revisi : 10 Juli 2015

ABSTRAK

Perguruan tinggi merupakan wahana yang diharapkan mampu menghasilkan sumber daya manusia yang unggul melalui penyelenggaraan pendidikan. Oleh karena itu, setiap perguruan tinggi perlu melakukan evaluasi atas setiap pelaksanaan proses pendidikan dan pengajaran yang dilakukan oleh dosen sebagai bagian tugas dan kewajiban dalam Tri Dharma perguruan tinggi. Pemetaan dosen berdasarkan kinerja dalam melaksanakan proses pendidikan dan pengajaran penting dilakukan untuk mengetahui seberapa banyak dosen yang telah melaksanakan tugas pendidikan dan pengajaran dengan mutu yang baik. Sekaligus mempertimbangkan program-program perbaikan dalam rangka membina para dosen dengan kualitas kinerja kurang baik. 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 algoritmanya yang sederhana dan mudah diimplementasikan. Penelitian ini memadukan antara metode PCA dengan *K-Means*. Metode PCA digunakan untuk mereduksi dimensi *dataset* sebelum di-*cluster* menggunakan *K-Means*. *Dataset* yang digunakan adalah data kinerja dosen Program Studi Teknik Elektro Universitas Panca Marga Probolinggo yang terdiri dari Inisial dosen, nilai aspek pedagogik, nilai aspek profesional, nilai aspek kepribadian dan nilai aspek sosial. Metode pengukuran *Davies-Bouldin (DB) Index* digunakan untuk mengukur validitas *cluster* yang dihasilkan. Nilai *DB Index* terkecil dihasilkan oleh algoritma *K-Means* yang diawali dengan reduksi dimensi *dataset* menggunakan PCA.

Kata kunci: Kinerja dosen, *K-Means*, *cluster*, *Principal Component Analysis*

ABSTRACT

The college is a vehicle that is expected to produce a superior human resources through education. Therefore, every university needs to conduct an evaluation of each implementation of the process of education and teaching conducted by a lecturer as part of the duties and obligations of the Tri Dharma. Mapping lecturer based on performance in carrying out the process of education and teaching is important to know how many lectures who have been carrying out educational and teaching duties in good qualities. While considering programs adapted in order to foster improvement of the quality of lecturers with less good performance. 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 advantages that the algorithm is simple and easy to implement. This study combines the PCA with K-Means. PCA is used to reduce the dimensions of the dataset before clustered using K-Means. The dataset used is a lecturer performance data of Electrical Engineering Program consisting of lecture's initial, pedagogic aspect value, the value of the professional aspects, value aspects of the personality and values of social aspects. The measurement method Davies-Bouldin (DB) Index is used to measure the validity of the resulting cluster. DB Value Index produced by the smallest K-Means algorithm that begins with dimension reduction using PCA dataset.

Keywords: Lecturer performance, *K-Means*, *cluster*, *Principal Component Analysis*

PENDAHULUAN

Perguruan tinggi merupakan wahana yang diharapkan mampu menghasilkan sumber daya manusia yang unggul melalui penyelenggaraan pendidikan. Pendidikan sebagai sebuah proses, terutama mengenal adanya *raw input* dan *instrumental input*. *Raw input* adalah peserta didik, sedangkan *instrumental input* terdiri dari : gedung, perpustakaan, pedoman akademik, dosen, kurikulum, metode, dan lain-lain. Kedua *raw input* dan *instrumental input* masuk dalam proses yang pelaksanaannya membutuhkan waktu setidaknya 8 semester. Ketiga, *output* (hasil didik) yang sesuai dengan kriteria institusi dan siap untuk masuk dalam persaingan sumber daya manusia [1].

Dosen merupakan instrumen yang sangat menentukan keberhasilan proses pendidikan, karena dari dosenlah perpindahan ilmu dilakukan kepada peserta didik. Perguruan tinggi yang memiliki tenaga dosen berkualitas akan banyak diminati oleh masyarakat [1]. Oleh karena itu, setiap perguruan tinggi perlu melakukan evaluasi atas setiap pelaksanaan proses pendidikan dan pengajaran yang dilakukan oleh dosen sebagai bagian tugas dan kewajiban dalam Tri Dharma perguruan tinggi. Pemetaan dosen berdasarkan kinerja dalam melaksanakan proses pendidikan dan pengajaran penting dilakukan untuk mengetahui seberapa banyak dosen yang telah melaksanakan tugas pendidikan dan pengajaran dengan mutu yang baik. Sekaligus mempertimbangkan proram-program perbaikan dalam rangka membina para dosen dengan kualitas kinerja kurang baik. 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 algoritmanya yang sederhana dan mudah diimplementasikan [2]. *K-Means* juga mempunyai kemampuan untuk mengelompokkan data dalam jumlah yang cukup besar dengan waktu komputasi yang relatif cepat dan efisien [3].

K-Means juga memiliki kelemahan yang diakibatkan oleh 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*. Selain itu, banyak algoritma klasifikasi dan *clustering* termasuk *K-Means* yang dihadapkan pada persoalan *curse of dimensionality* saat

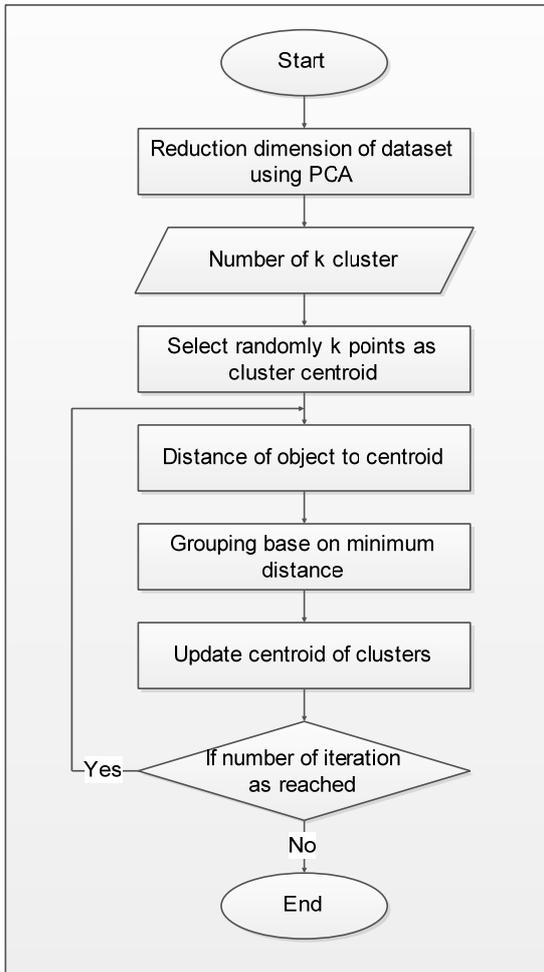
dihadapkan pada persoalan data dengan fitur yang besar. Algoritma klasifikasi dan *clustering* menjadi bermasalah pada data dengan dimensi tinggi berupa menurunnya akurasi klasifikasi dan kualitas *cluster* yang jelek dan juga berpengaruh pada waktu komputasi yang lama. Permasalahan ini dapat dapat diatasi dengan melakukan reduksi dimensi.

Banyak keuntungan yang didapat dengan melakukan reduksi dimensi. Banyak algoritma *data mining* yang bekerja dengan baik jika dimensi (jumlah atribut/fitur pada data) lebih rendah. Alasannya adalah reduksi dimensi dapat menghilangkan fitur yang tidak relevan, dan mengurangi *noise*, selain itu juga mengurangi *curse of dimensionality*. Banyak metode yang bisa digunakan untuk reduksi dimensi. Salah satunya adalah metode *Principal Component Analysis* (PCA). PCA dapat mereduksi dimensi data yang tinggi menjadi dimensi data yang lebih rendah dengan resiko kehilangan informasi yang sangat kecil [6].

Penelitian ini mengusulkan metode PCA untuk mereduksi dimensi data untuk optimasi *cluster* pada algoritma *K-Means*.

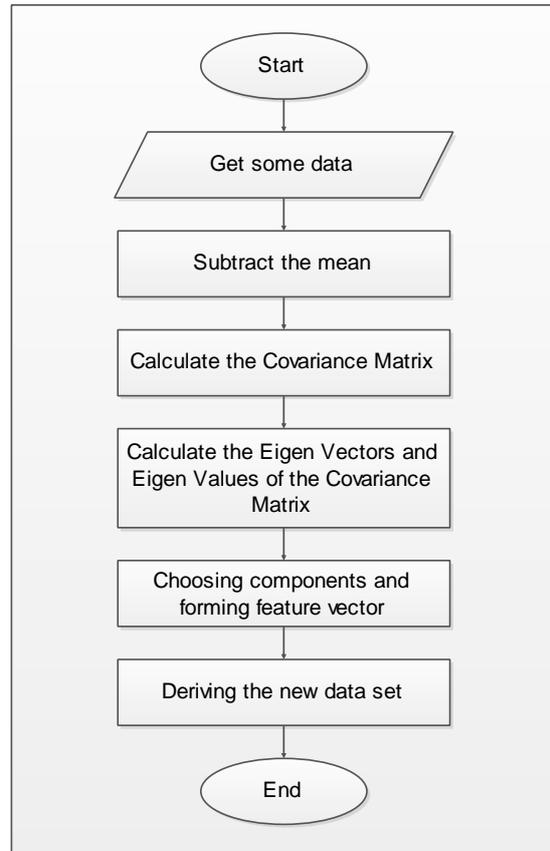
METODE PENELITIAN

Penelitian ini mengkombinasikan algoritma *K-Means* tradisional dengan metode PCA. Sebelum proses *clustering* dilakukan, reduksi dimensi *dataset* dilakukan menggunakan 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 1 menunjukkan algoritma *K-Means* yang dikombinasikan dengan metode PCA.



Gambar 1. Kombinasi Algoritma *K-Means* dengan PCA

Reduksi dimensi *dataset* dilakukan dengan menggunakan metode PCA. Metode PCA dilakukan sesuai dengan tahapan-tahapan yang digambarkan pada Gambar 2.



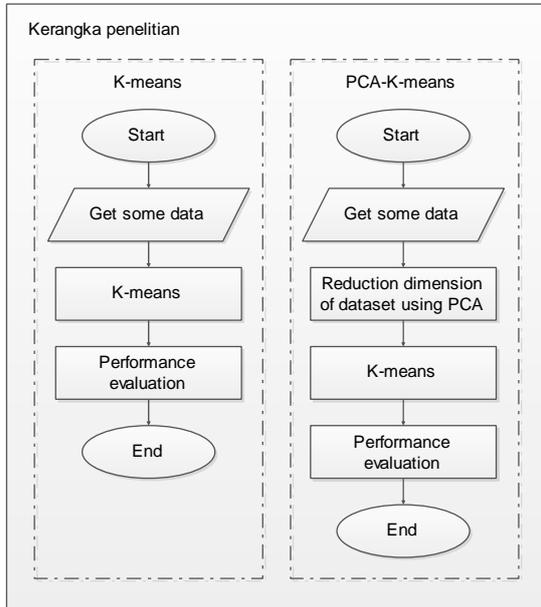
Gambar 2. Algoritma metode PCA

Setelah reduksi dimensi *dataset* dilakukan dengan metode PCA, langkah selanjutnya adalah *clustering* dengan *K-Means*. Berikut tahapan algoritma *K-Means* tradisional [2]:

1. Membuat partisi sejumlah k dari segmentasi yang akan dibentuk
2. Pilih secara acak k point untuk dijadikan pusat *cluster*
3. Menghitung jarak data yang lain dengan pusat *cluster*
4. Mengisi setiap objek dalam *dataset* ke dalam segmen terdekat
5. Kalkulasi ulang setiap segmentasi yang terbentuk
6. Ulangi langkah hingga data di dalam segmentasi tidak berubah

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 (PCA-*K-*

Means). Kerangka penelitian ini ditunjukkan dalam gambar 3.



Gambar 3. Kerangka penelitian

Penelitian ini dilakukan dengan menggunakan alat bantu Rapid Miner.

HASIL DAN PEMBAHASAN

Penelitian ini dilakukan melalui beberapa tahapan meliputi : pengumpulan data dan pemilihan atribut, tahap reduksi dimensi *dataset*, tahap *clustering* dan pengujian performansi algoritma.

Pengumpulan data dan pemilihan atribut

Data dalam penelitian ini adalah data hasil evaluasi kinerja dosen Program Studi Teknik Elektro Fakultas Teknik Universitas Panca Marga Probolinggo Tahun Akademik 2014/2015. Data terdiri dari beberapa atribut meliputi : nama dosen/ inisial, nilai untuk aspek pedagogik (a), nilai untuk aspek profesional (b), nilai untuk aspek kepribadian (c), nilai untuk aspek sosial (d) dan jumlah mata kuliah yang diampu (e).

Tabel 1. Data kinerja dosen bidang pendidikan dan pengajaran

No	Inisial	(a)	(b)	(c)	(d)	(e)
1	IW	6,53	6,55	6,58	6,78	1
2	NH	6,33	6,14	6,20	6,07	2

Lanjutan Tabel 1...

No	Inisial	(a)	(b)	(c)	(d)	(e)
3	DP	6,13	6,03	6,23	6,24	1
4	AS	6,09	5,77	6,13	6,24	1
5	TR	5,76	5,25	6,73	6,20	1
6	HW	5,96	5,63	6,10	6,20	1
7	AI	5,83	5,56	6,17	6,10	3
8	ASR	5,75	5,75	5,95	5,91	1
9	KM	5,60	5,56	6,03	5,96	1
10	DPK	5,73	5,51	5,83	5,86	3
11	DMD	5,57	5,50	6,05	5,71	2
12	IZ	5,34	5,28	5,77	5,39	2
13	MI	5,33	5,13	5,63	5,68	3
14	AB	4,77	5,40	5,78	5,20	1
15	AAR	5,42	5,00	5,55	5,00	1
16	TA	4,81	5,14	5,37	5,18	2
17	AH	4,81	4,92	5,33	5,13	1
18	DIH	5,06	4,93	5,00	5,13	1
19	MO	4,83	4,77	5,01	4,82	1
20	NLM	4,47	4,58	4,93	4,72	1
21	ANS	4,47	4,02	4,66	4,90	3
22	DA	4,26	4,18	4,60	4,08	1
23	RY	3,60	3,98	4,87	4,50	1
24	IM	4,13	3,75	4,25	4,14	3
25	EK	3,91	3,84	3,95	4,24	3
26	AIS	2,37	2,14	2,17	2,02	1

Tidak semua atribut digunakan dalam penelitian ini, atribut yang digunakan adalah (a), (b), (c) dan (d). Tahapan berikutnya adalah tahap reduksi dimensi *dataset* dengan metode PCA.

Reduksi dimensi *dataset* dengan PCA

Pada tahap ini dilakukan reduksi dimensi *dataset* menggunakan PCA. Penelitian ini mereduksi dimensi *dataset* menjadi data 1 dimensi (saat memutuskan hanya memilih 1 *eigen vector*) dan 2 dimensi (saat memutuskan untuk memilih 2 *eigen vector*). Hasil reduksi dimensi *dataset* ditunjukkan pada Tabel 2.

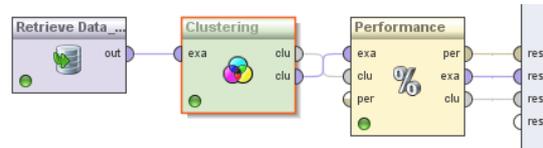
Tabel 2. Reduksi dimensi *dataset* menjadi data 1 dimensi dan data 2 dimensi menggunakan PCA

Reduksi menjadi 1 dimensi			Reduksi menjadi 2 dimensi			
No	Inisial	PC1	No	Inisial	PC1	PC2
1	IW	2.809	1	IW	2.809	-0.237
2	NH	1.950	2	NH	1.950	-0.397
3	DP	1.901	3	DP	1.901	-0.190
4	AS	1.703	4	AS	1.703	-0.156
5	TR	1.570	5	TR	1.570	0.619
6	HW	1.531	6	HW	1.531	-0.053
7	AI	1.423	7	AI	1.423	0.074
8	ASR	1.264	8	ASR	1.264	-0.128
9	KM	1.166	9	KM	1.166	0.094
10	DPK	1.058	10	DPK	1.058	-0.135
11	DMD	1.004	11	DMD	1.004	0.087
12	IZ	0.473	12	IZ	0.473	0.037
13	MI	0.479	13	MI	0.479	0.054
14	AB	0.157	14	AB	0.157	0.328
15	AAR	0.067	15	AAR	0.067	-0.163
16	TA	-0.166	16	TA	-0.166	0.094
17	AH	0.313	17	AH	0.313	0.123
18	DIH	-0.354	18	DIH	-0.354	-0.258
19	MO	-0.698	19	MO	-0.698	-0.132
20	NLM	-1.066	20	NLM	-1.066	0.082
21	ANS	1.381	21	ANS	1.381	0.094
22	DA	-1.858	22	DA	-1.858	-0.33
23	RY	-1.932	23	RY	-1.932	0.713
24	IM	-2.274	24	IM	-2.274	-0.052
25	EK	-2.438	25	EK	-2.438	-0.126
26	AIS	-6.074	26	AIS	-6.074	-0.342

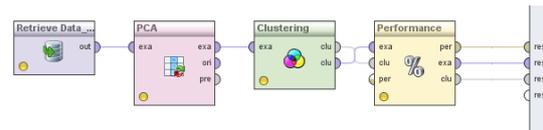
Setelah dilakukan reduksi dimensi *dataset* menggunakan PCA, tahapan berikutnya adalah tahap *clustering*.

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 4 dan Gambar 5.



Gambar 4. Pemodelan *K-Means* dengan *Rapid Miner*



Gambar 5. Pemodelan *PCA-K-Means* dengan *Rapid Miner*

Dilakukan beberapa kali percobaan terhadap model yang dibangun. Model pertama adalah melakukan *clustering* terhadap *dataset* tanpa melakukan reduksi dimensi (*K-Means* tradisional). Jumlah *cluster* ditentukan mulai dari 2, 3, 4, dan 5. Hasil segmentasi yang terbentuk akan dievaluasi menggunakan *Davies-Bouldin (DB) Index*. *Davies-Bouldin Index* adalah fungsi rasio dari jumlah antara *cluster scatter* sampai dengan *cluster separation*. *Davies-Bouldin Index* merupakan metode validasi *cluster* dari hasil *clustering*. Pendekatan pengukuran DBI yaitu memaksimalkan jarak *inter cluster* serta meminimalkan jarak *intra cluster*.

Pengujian performansi algoritma

Tabel 3 merupakan nilai *DB Index* berdasarkan hasil percobaan untuk setiap model menggunakan *Rapid Miner*.

Tabel 3. Nilai *DB Index* hasil percobaan

Model <i>clustering</i>	Jumlah <i>cluster</i> (k)			
	2	3	4	5
<i>K-Means</i>	0.846	0.673	0.583	0.579
<i>PCA-K-Means</i> (<i>dataset</i> 1 dimensi)	0.551	0.398	0.346	0.462
<i>PCA-K-Means</i> (<i>dataset</i> 2 dimensi)	0.585	0.426	0.420	0.528

Dari hasil percobaan yang dilakukan, algoritma *K-Means* dengan PCA dapat menghasilkan kualitas *cluster* yang lebih baik dibandingkan dengan *K-Means* tradisional. Akhirnya hasil evaluasi *cluster* menunjukkan bahwa algoritma *K-Means* yang diawali

dengan reduksi dimensi *dataset* menjadi 1 dimensi menggunakan PCA lebih optimal dengan nilai *DB Index* sebesar 0.346. Jumlah *cluster* yang terbentuk adalah 4 dengan segmentasi sebagai berikut :

1. *Cluster 0* : 11 item
2. *Cluster 1* : 8 item
3. *Cluster 2* : 6 item
4. *Cluster 3* : 1 item

Cluster 0 merupakan kumpulan dosen yang memiliki nilai baik untuk semua atribut yang bisa dikelompokkan sebagai dosen yang memiliki kinerja sangat baik. *Cluster 1* merupakan kelompok dosen dengan kinerja cukup baik, *Cluster 2* merupakan kelompok dosen dengan kinerja kurang baik dan yang terakhir, *Cluster 3* adalah dosen dengan kinerja tidak baik. Tabel 4 berisi hasil *cluster* dosen dengan PCA-K-Means.

Tabel 4. Hasil Cluster dengan PCA-K-Means

Cluster	Jumlah item	Inisial dosen
Cluster 0	11	IW, NH, DP, AS, TR, HW, AI, ASR, KM, DPK, DMD
Cluster 1	8	IZ, MI, AB, AAR, TA, AH, DIH, MO
Cluster 2	6	NLM, ANS, DA, RY, IM, EK
Cluster 3	1	AIS

Berdasarkan penyajian yang diberikan pada bagian pembahasan ini, metode reduksi dimensi *dataset* menggunakan PCA dapat digunakan untuk optimasi *cluster* pada algoritma K-Means.

SIMPULAN

Secara umum, metode PCA dapat digunakan untuk mereduksi dimensi *dataset* tanpa harus kehilangan banyak informasi. Dari penelitian yang dilakukan, reduksi dimensi *dataset* menggunakan metode PCA terbukti dapat meningkatkan kualitas *cluster* yang dihasilkan oleh algoritma K-Means. Peningkatan kualitas model dapat dilihat dari peningkatan akurasi yang cukup signifikan. Hasil pengukuran validitas *cluster* dengan *Davies-Bouldin Index* menunjukkan PCA-K-Means menghasilkan nilai *DB Index* paling kecil dibandingkan dengan K-Means tradisional.

Meskipun model yang diusulkan sudah memberikan hasil yang lebih baik, namun untuk penelitian selanjutnya dapat menerapkan PCA-K-Means untuk *dataset* yang lebih beragam dan berbeda.

DAFTAR PUSTAKA

- [1] Asmawi, Rosul, M., 2005. *Strategi Meningkatkan Lulusan Bermutu di Perguruan Tinggi*. Makara, Sosial Humaniora, Vol. 9, No. 2, Desember 2005
- [2] Widiarina, Wahono, Satria, Romi, 2015. *Algoritma Cluster Dinamik untuk Optimasi Cluster pada Algoritma K-Means dalam Pemetaan Nasabah Potensial*. Journal of Intelligent Systems, Vol. 1 No. 1, February 2015, ISSN 2356-3982
- [3] Tahta, Alfina, Santosa, Budi, Barakbah, Ridho, Ali, 2012. *Analisa Perbandingan Metode Hierarchical Clustering, K-Means dan Gabungan Keduanya dalam Cluster Data (Studi kasus: Problem Kerja Praktek Jurusan Teknik Industri ITS)*. Jurnal Teknik ITS Vol. 1, September 2012, ISSN 2301-9271
- [4] Prasetyo, Eko, 2014. *Reduksi Dimensi Set Data dengan DRC pada Metode Klasifikasi SVM dengan Upaya Penambahan Komponen Ketiga*. Prosiding Snatif Ke-1. ISBN: 978-602-1180-0404
- [5] Ramadhani, Dias, Rima, 2014. *Data Mining Menggunakan Algoritma K-Means Clustering untuk Menentukan Strategi Promosi Universitas Dian Nuswantoro*. Jurusan Sistem Informasi, Fakultas Ilmu Komputer, Universitas Dian Nuswantoro
- [6] Smith, I., Lindsay, 2002. *A Tutorial on Principal Component Analysis*.