

Algoritma K-Means Dalam Mengidentifikasi Perkerjaan Berdasarkan Latar Belakang Pendidikan di Tingkat Rukun Warga.

Erza Sofian

Program Studi Sistem Informasi, STIMIK ESQ
Jl. TB Simatupang Kaveling 1, Cilandaki, Jakarta Selatan – 12560
Email: erza.sofian@esqbs.ac.id

Abstract: *This research is meant to identify level of unemployment based on education background that is hold each community that has graduated from university or college. The total number of graduates that graduated in each university is significant compare to available work fields, that resulted in high level of competition which made companies also set high standard in recruitment. Algorithm K-means with clustering technique can identify total population for graduates according to their root of education, age, job and available work fields. This data is produced qualitative gradually to fulfill data research needs. Therefore, it is expected that from this research can come as recommendation to each candidate or potential students in taking steps for the future and give training in accordance with their study field that graduate student or potential graduate student holds.*

Keywords: *Identification, University/College, K-means, Qualitative, Recommendation.*

Abstrak: Dalam penelitian ini dimaksudkan untuk mengidentifikasi tingkat pengangguran berdasarkan latar pendidikan yang di sandang setiap masyarakat yang telah menyelesaikan pendidikan perguruan tinggi. Jumlah lulusan yang di hasilkan dari setiap perguruan tinggi cukup signifikan di bandingkan dengan jumlah lapangan pekerjaan yang tersedia, sehingga tingkat kompetitif cukup tinggi dalam persaingan, sehingga perusahaan-perusahaan membuat standard penerimaan yang cukup tinggi pula. Algoritma K-means dengan teknik clustering dapat mengidentifikasi jumlah populasi lulusan perguruan tinggi sesuai dengan rumpun ilmunya, umur, pekerjaan dan lapangan pekerjaan yang tersedia. Data ini di hasilkan secara kualitatif yang di lakukan secara bertahap untuk memenuhi kebutuhan data penelitian. Oleh karenin hasil dari penelitian di harapkan menghasilkan suatu rekomendasi pada setiap kandidat atau calon mahasiswa dalam mengambil langkah masa depannya serta memberikan pelatihan-pelatihan keterampilan sesuai dengan bidang ilmu yang di miliki bagi lulusan atau bagi calon lulusan perguruan tinggi.

Kata kunci: Identifikasi, perguruan tinggi, K-means, kualitatif, rekomendasi

1. PENDAHULUAN

Berdasarkan data Badan Pusat Statistik(BPS) tahun 2014, struktur lapangan pekerjaan hingga februari 2014 tidak mengalami perubahan, dimana sector pertanian, perdagangan, jasa kemasyarakatan, dan sector industry secara berurutan masih menjadi penyumbang terbesar penyerapan tenaga kerja di indonesia. Jika di bandingkan dengan keadaan pada bulan februari 2013, jumlah penduduk yang bekerja mengalami kenaikan pada hampir semua sektor terutama di Sektor Jasa Kemasyarakatan sebanyak 640 ribu orang (3,59%), sektor Perdagangan sebanyak 450 ribu orang (1,77%), serta Sektor Industri sebanyak 390 ribu orang (2,60%), sedangkan yang mengalami penurunan hanya sektor pertanian sebanyak 280 ribu orang (0,68%).

Sedangkan Penyerapan tenaga kerja berdasarkan pendidikan pada bulan ferbruari 2014 masih di dominasi oleh penduduk bekerja berpendidikan rendah yaitu SD ke bawah sebanyak 55,3 juta orang (46,80%) dan sekolah menengah pertama sebanyak 21,1 juta (17,82%) berpendidikan Diploma dan sebanyak 8,8 juta orang (7,49%) berpendidikan Universitas.

Perbaikan kualitas penduduk yang bekerja di tunjukan oleh kecendrungan menurunnya penduduk bekerja berpendidikan rendah (SMP kebawah) dan meningkatnya penduduk bekerja berpendidikan tinggi (Diploma dan Univertsitas). Dalam setahun terakhir, penduduk bekerja perpendidikan rendah menurun dari sebanyak 76,8 juta orang (66 %) pada Februari 2013 menjadi 76,4 juta orang (64,63%).

Melihat data di atas yang terjadi pada tahun 2013-2014 di mana banyak beberapa industri yang tutup dalam usahanya, hal ini

dapat menimbulkan suatu prediksi untuk tingkat pengangguran yang terjadi pada di bawah tahun 2014 dengan mangmbil data BPS dari jumlah penurunan tingkat pengangguran, dari informasi tersebut dapat dibandingkan dengan jumlah sektor industri business yang ada sehingga dapat disimpulkan jumlah pengangguran berkurang pada tingkat pendidikan tinggi dikarenakan mereka mengisi lapangan pekerjaan untuk tingkat pendidikan rendah.

Rumusan masalah penelitian ini dilatarbelakangi pemikiran dari beberapa masalah social yang sering terjadi di beberapa wilayah Jakarta, masalah ini yang sangat berdampak pada masyarakat lainnya antara lain meningkatnya tindak kriminal sampai perselisihan antara beberapa kelompok. Penyebab yang menimbulkan tingkat pengangguran pada suatu wilayah yang memiliki latar belakang pendidikan tinggi serta kesediaan lapangan pekerjaan yang tidak sesuai dengan bidang keilmuannya.

1.1. Identifikasi masalah

Tujuan dari penelitian ini adalah pengaplikasian metode clustering K-means dalam mengelompokan pengangguran berdasarkan latar belakang pendidikan dan lapangan pekerjaan yang tersedia, sehingga dapat mengetahui:

1. Jumlah pengangguran pada suatu wilayah berdasarkan tingkat latar belakang pendidikan.
2. Dampak pengaruh social pada wilayah tersebut.
3. Dampak pengaruh penghasilan masyarakat pada wilayah terebut yang mempengaruhi perkembangan wailayahnya.

1.2. Rumusan Masalah

Sesuai dengan hubungan yang telah dibatasi maka dibangun rumusan masalah sebagai berikut:

1. Pengaruh dan dampak dari tingkat tingkat pendidikan terhadap lapangan pekerjaan?
2. Keterkaitan lapangan pekerjaan yang tersedia dengan latar belakang pendidikan dapat memberikan solusi dalam mengatasi atau menekan tingkat pengangguran bisa lebih kecil.

2. TINJAUAN PUSTAKA

Farid Alghofar (2007), Fakultas Ekonomi Universitas Dipenogoro, dalam penelitiannya berjudul "Analisa Tingkat Pengangguran di Indonesia tahun 1980-2007" dalam salah satu point tujuan dari penelitiannya adalah menganalisa hubungan antara kenaikan pertumbuhan ekonomi yang lambat dengan kenaikan angka pengangguran. Point ini merupakan salah satu referensi atau rujukan salah satu penyebab tingkat pengangguran di suatu wilayah yang akan di gunakan sebagai data pembandingan.

Sukirno (2004), dalam tulisannya "Terdapat kaitan erat diantara masalah kejahatan dan masalah pengangguran, yaitu semakin tinggi pengangguran semakin tinggi kasus kejahatan. Dengan demikian usaha mengatasi pengangguran secara tak langsung menyebabkan pengurangan dalam kejahatan" (Makro ekonomi teori pengantar). Hal ini dapat dianalisa hubungan kasus kejahatan yang terjadi pada suatu wilayah atau daerah di sebabkan oleh tingkat pengangguran yang tinggi, hal ini mejadi kajian yang harus di selesaikan untuk mencari solusi dalam mengurangi tingkat pengangguran.

3. METODE PENELITIAN

3.1. K-means

Metode K-means suatu metode clustersing dengang menggunakan pendekatan pengelompokan clustering (Partitional Clustering) yang dilakukan secara pengulangan dalam menghasilkan jumlah cluster, K, kemudian proses ini dilanjutkan dengan menghubungkan tiap tiap titik cluster (centroid). Penentuan nilai centroid dilakukan secara acak (random) sebagai

$$V_{ij} = \frac{1}{N_i} \sum_{k=0}^{N_i} X_{kj}$$

awal iterasi, selanjutnya dalam menentukan nilai centroid pada tahap berikutnya dapat di formulasikan,

X_{kj} merupakan nilai data ke- k yang ada dalam cluster tersebut untuk variabel ke- j

V_{ij} , merupakan centroid / rata-rata cluster ke- i untuk variabel j .

N_i merupakan Jumlah data yang menjadi anggota cluster ke - i

i, k merupakan indeks dari cluster

j merupakan indeks dari variabel

Dalam penghitugan jarak pada data yang ada terhadap pusat data (centroid) pada cluster yang sudah di tentukan dapat di gunakan beberapa rumusan yang tersedia, dalam penelitian ini digunakan rumus :

Jarak dua titik x dan y menurut Euclidean dirumuskan sebagai:

$$d(x,y) = || x - y ||^2 = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

Penggunaan rumus Euclidean akan di gunakan perhitungan jarak dalam distance space, yaitu merupakan jarak pendek yang bisa di dapatkan antara dua titik yang di perhitungkan.

3.2. Transformasi Data

Agar data dapat di proses dengan menggunakan Metode K-mean cluster, terlebih dahulu data harus di klasifikasikan seperti data yang berjenis data nominal seperti wilayah dan jumlah penduduk serta tingkat pendidikan dan jurusan harus diinisialisasikan dalam bentuk angka:

1. Pengelompokan elemen yang dibagi beberapa beberapa bagian, seperti kecamatan yang terbagi beberapa wilayah kelurahan.

Rukun Warga terdiri dari beberapa Rukun Tetangga, Rt.001 sampai dengan Rt.017

2. Kemudian wilayah-wilayah di urutkan berdasarkan jumlah populasinya.
3. Kemudian di berikan penomoran atau angka secara terurut, sebagai inisialisasi pada wilayah dengan jumlah populasi penduduk yang tinggi – rendah.

3.2. Konsep Jarak.

3.2.1. Jarak Dua Titik

Untuk mengukur jarak dua titik x dan y, ($d(x,y)$), kita bisa menggunakan beberapa konsep jarak. Ukuran jarak harus memenuhi syarat-syarat sebagai berikut:

1. $d(x,y) \geq 0$ (non-negatif). Tidak ada jarak yang mempunyai nilai negatif.
2. $d(x,y)=0$ jika dan hanya jika $x=y$ (identity of indiscernibles) jarak antara suatu obyek atau titik dengan objek atau titik itu sendiri adalah nol.
3. $d(x,y) = d(y,x)$ (simetri) . Jarak dari x ke y adalah sama dengan jarak dari y ke x.
4. $d(x,z) \leq d(x,y) + d(y,z)$ (ketidaksamaan segitiga).

Beberapa macam jarak yang sering digunakan dalam literature machine learning / data mining adalah:

3.2.2. Jarak Euclidean

Jarak dua titik x dan y menurut Euclidean dirumuskan sebagai:

$$d(x,y) = || x - y ||^2 = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

3.2.3. Jarak antara 2 (dua) klaster

Ada beberapa cara dalam menentukan jarak antara dua klaster, klaster 1 dan klaster 2, antara lain digunakan beberapa cara:

1. Jarak maksimum antara elemen dalam klaster (*complete linkage clustering*).

Dengan cara ini jarak antara dua klaster di definisikan sebagai

$$d(A,B) = \max_{x \in A, y \in B} \{ S_{xy} \}$$

Dimana S_{xy} adalah jarak dua data x dan y masing – masing dari klaster A dan B

2. Jarak Minimum antara elemen dari setiap klaster (*single linkage clustering*).

Dengan cara ini jarak antara kedua klaster di definisikan sebagai

$$d(A,B) = \min_{x \in A, y \in B} \{ S_{xy} \}$$

3. Rata-rata jarak antara elemen dari setiap klaster (*average linkage clustering*)

$$d(A,B) = \frac{1}{n_A n_B} \sum_{x \in A} \sum_{y \in B} s\{x,y\}$$

Dimana n_A dan n_B masing masing adalah banyaknya data dalam set A dan B

4. Centroid Linkage

Dengan cara ini jarak antara ke dua klaster di definisikan sebagai

$$d(A,B) = s(\bar{x}, \bar{y})$$

$$\text{dimana } \bar{x} = \frac{1}{n_A} \sum_{x \in A} x,$$

$$\bar{y} = \frac{1}{n_B} \sum_{y \in B} y,$$

4. IMPLEMENTASI

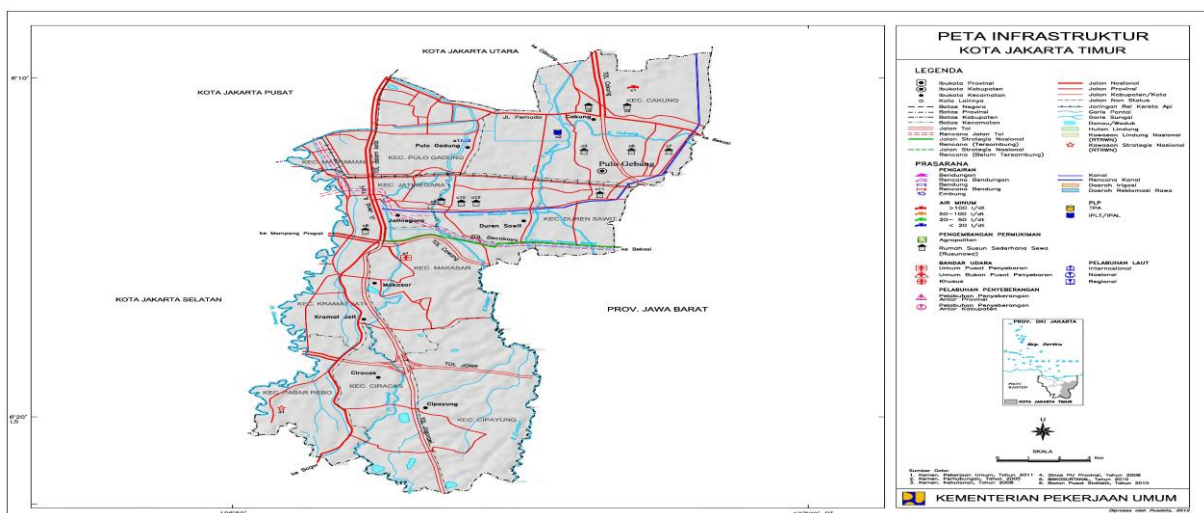
Pada penelitian ini dimulai dari melakukan pengamatan secara langsung pada

lingkungan masyarakat berada. Pengamatan secara langsung ini dilakukan melalui wawancara terhadap masyarakat yang mengalami kendala terhadap kegiatan yang tidak sesuai dengan latar belakang pendidikan yang dimiliki. Wawancara ini bertujuan untuk mengetahui gambaran permasalahan yang ada di lingkungan masyarakat tersebut. Berdasarkan hasil wawancara yang di lakukan sehingga dapat di tetapkan rumusan masalah yang ada di lingkungan masyarakat tersebut.

Data-data yang digunakan pada penelitian ini didapat dari pengambilan data secara langsung dengan memberikan form pengisian data pada masyarakat, sehingga kita mendapatkan data yang terkumpul dari isian yang kita berikan. Data-data ini berisi

data pribadi dari masyarakat yang kita kumpulkan. Kemudian data kita klasifikasikan atau kita kelompokkan berdasarkan tingkat Nama, tahun lahir, pendidikan Akhir, Jenis Kelamin dan Pekerjaan sebelumnya dan pekerjaan sekarang.

Data-data yang di telah di dapatkan pada tahap pengumpulan data kemudian di lakukan transformasi pada data data yang berjenis data nominal seperti pendidikan akhir dan pekerjaan. Data data yang berjenis data nominal ini diinisialisasikan ke dalam bentuk angka melalui beberapa langkah agar data – data yang berjenis data nominal ini dapat di olah dengan menggunakan **Algoritma K-means Clustering**.



Gambar.1 Peta Infrastruktur Jakarta Timur.

5. TRANSFORMASI DATA

Tabel.1 Data Penduduk

Nama	Jenis_Kelamin	Tgl_Lahir	Pend_Akhir	Pekerjaan
Xxxxxxxx	P/L	1970	(S1, S2)	Apa ?

Untuk data di atas dapat di olah dengan menggunakan metode K-means Clustering, maka data berjenis data nominal seperti

pendidikan akhir dan pekerjaan diinisialisasikan terlebih dahulu dalam bentuk angka.

Untuk melakukan inisialisasi Pendidikan Akhir di lakukan dengan cara,

STRATA 1.....

STRATA 2.....

6. KEBUTUHAN DATA

6.1. Menentukan Variable

Dari identifikasi masalah yang ada, penentuan data yang di perlukan untuk proses clustering yaitu data nilai yang di gunakan dalam proses penelitian di peroleh dari

Tabel 2a. Rumpun Ilmu Teknik

Rumpun Ilmu Teknik	
Sub Rumpun Ilmu Teknik Elektro dan Informatika	
Bidang Ilmu	Teknik Elektro Teknik Tenaga Elektrik Teknik Telekomunikasi Teknik Kendali Teknik Biomedika Teknik Informatika Ilmu Komputer Sistem Informasi Teknologi Informasi Teknik Perangkat Lunak DII

Tabel 2b. Rumpun Ilmu Ekonomi

Rumpun Ilmu Ekonomi
Sub Rumpun Ilmu Ekonomi dan Ilmu Manajemen

Bidang Ilmu	Akutansi Ekonomi Syariah Perbankan Perpajakan Manajemen Manajemen Informatika Pemasaran Manajemen Industri Manajemen Syariah DII
-------------	---

Tabel 2c. Rumpun Ilmu Sosial dan Humaniora

Rumpun Sosial dan Humaniora	
Sub Rumpun Ilmu Politik	
Bidang Ilmu	Kriminologi Hub International Ilmu Hukum Ilmu Pemerintahan Ketahanan Nasional Ilmu Kesejahteraan Sosial Sosiologi Ilmu Komunikasi Antropologi Kajian Wilayah Sejarah DII

6.2. Menentukan Sampel

Dalam Penelitian Penentuan sampel merupakan kebutuhan utama untuk memperoleh data yang menggambarkan

realitas yang akan di jadikan fokus dalam

Vr	Bidang Ilmu	Rumpun Ilmu
R1	Teknik Informatika	Teknik
R2	Ilmu Komputer	
R3	Sistem Informasi	
R4	Teknologi Informasi	
R5	Manajemen	Ekonomi
R6	Manajemen Informatika	
R7	Pemasaran	
R8	Manajemen Industri	
R9	Ketahanan Nasional	Humaniora
R10	Ilmu Kesejahteraan Sosial	
R11	Sosiologi	
R12	Ilmu Komunikasi	

penelitian.

Rumusan Solvin :

$$n = \frac{N}{1 + Ne^2}$$

n = Jumlah sample

N= Jumlah Populasi

e² = Toleransi ketidakteelitian

6.3. Inisialisasi Data Numerik

Dalam Metode clustering k-means menggunakan tipe data berupa numerik . Kemudian data yang di ubah di kelompokkan berdasarkan rumpun ilmu yang sesuai dengan konsentrasi sub rumpun ilmu, yang di ambil berdasarkan data dari lapangan yang terkumpul. Kemudian data tersebut di Klasifikan kembali untuk menentukan variabel yang akan di gunakan. Variabel ini di di lakuaan terurut sesuai order dari tiap subrumpun ilmu yang ada. Kemudian di buat nilai rata-rata nya berdasarkan untuk

di bentuk beberapa group di dalam tulian ini di ambil sampel untuk 3 group rumpun ilmu sesuai denga subrumpun ilmu tersebut.

6.4. Klasifikasi Variable

X_{SH} = Sosial dan Humaniora

X_{IT} = Ilmu Teknik

X_{IE} = Ilmu Ekonomi

$$X_{IT} = \frac{R1 + R2 + R3 + R4}{4}$$

(Penghasilan rata-rata per tahun Ilmu Teknik)

$$X_{SH} = \frac{R1 + R2 + R3 + R4}{4}$$

(Penghasilan rata-rata per tahun Ilmu Humaniora dan Sosial)

$$X_{IE} = \frac{R1 + R2 + R3 + R4}{4}$$

(Penghasilan rata-rata per tahun Ilmu Ekonomi)

$$X_{IT} = \frac{15 + 12 + 10 + 12}{4}$$

(Penghasilan rata-rata dalam juta)

Rata rata : X_{IT} = 12.5 ; X_{SH} = 10.05 : X_{IE} = 8.75 (hanya nilai estimasi)

6.5. Data Populasi Wilayah Rukun Warga

Data populasi untuk wilayah RW 03 di ambil berdasarkan data yang ada dari tiap-tiap pencatatan dan konformasi dari kantor sekretariat Rukun Warga. Data populasi ini di masukan berdasarkan per-kepala yang ada dari rumah tinggal.

Tabel 6.5a. Populasi 1

No	Wilayah	Volume/Populasi	Identitas
1	Rt.001	44 Jiwa	17
2	Rt.002	139 Jiwa	9

3	Rt.003	105 Jiwa	14
4	Rt.004	124 Jiwa	13
5	Rt.005	197 Jiwa	2
6	Rt.006	155 Jiwa	7
7	Rt.007	128 Jiwa	12
8	Rt.008	70 Jiwa	16
9	Rt.009	160 Jiwa	5
10	Rt.010	150 Jiwa	10
11	Rt.011	136 Jiwa	11
12	Rt.012	215 Jiwa	1
13	Rt.013	188 Jiwa	3
14	Rt.014	105 Jiwa	15
15	Rt.015	135 Jiwa	8
16	Rt.016	167 Jiwa	4
17	Rt.017	159 Jiwa	6
	Total	2385 Jiwa	

Tabel 6.5a. Populasi 2

Wilayah	XIT	XSH	XIE	Total	Dataset
Rt.001	1	1	4	6	1
Rt.002	6	4	6	16	2
Rt.003	4	1	4	9	3
Rt.004	4	3	9	16	4
Rt.005	10	3	12	25	5
Rt.006	4	3	9	16	6
Rt.007	5	1	6	12	7
Rt.008	4	1	5	10	8
Rt.009	8	3	9	20	9
Rt.010	15	6	7	28	10
Rt.011	5	3	10	18	11
Rt.012	8	4	9	20	12
Rt.013	8	4	12	24	13
Rt.014	10	2	8	20	14
Rt.015	16	4	10	30	15
Rt.016	20	5	15	40	16
Rt.017	8	2	25	35	17
	135	50	160	345	

6.6. Proses Sample

$$n = \frac{N}{1 + Ne^2}$$

$$n = \frac{2385}{1 + 2385 \times (0.05)^2}$$

$$n = \frac{2385}{1 + 2385 \times 0.0025}$$

$$n = \frac{2385}{6,9625}$$

$$n = 342$$

6.7. Proses Jarak (Euclidean Distance)

$$d(x,y) = ||x - y||^2 = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

$$I = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

$$II = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

$$III = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

$$I$$

$$= \sqrt{(12.5 - 10.5)^2 + (10.5 - 8.75)^2 + (8.75 - 8.75)^2}$$

$$I = 2.39$$

Dari perhitungan di atas sebagai perhitungan jarak data pusat cluster antara objek ke centroid dengan perhitungan Euclidian. Kemudian perhitungan selanjutnya untuk data set selanjutnya. Setelah data di dapat dari jarak objek ke centroid yitu bidang ilmu dengan centroid. Sehingga di dapatkan centroid akhir yaitu perhitungan centroid dengan dengan nilai rata-rata ini dari hasil perhitungan untuk mendapatkan update centroid tiap data ke -n sesuai dengan jumlah dari cluster cluster yang ada,

7. KESIMPULAN DAN SARAN

Dengan metode clustering k-means dapat melakukan proses penggalan informasi-informasi serta dapat mengelompokan data-data populasi penduduk dengan latar belakang pendidikan sarjana dalam bidang

ilmu berbeda-beda dengan lapangan pekerjaan yang berbeda-beda pula. Serta dapat pula mengelompokkan bidang ilmu yang di miliki dan hubungan dengan tingkat pengangguran serta lapangan pekerjaan yang tersedia.

Learning Tools and Techniques, Second Edition”, Morgan Kaufman, San Fransisco, 2005.

DAFTAR PUSTAKA

- [1] <http://www.ieee.org.ar>. Accomplishments & Future Directions. Diakses 18 february 2014.
- [2]<http://www.ieee.org.ar/downloads/Srivastava-tut-pres.pdf>.
- [3] <http://www.komputasi.lipi.go.id>
- [4] Anto Satriyo Nugroho , “Data mining dalam Bioinformatika: Menggaali Informasi Terpendam Dalam Lautan Data Biologi”, BPPT.
- [5] Larose, Daniel T, “Discovering Knowledge in Data: An Introduction to Data Mining”, John Willy and Son. Inc, 2005.
- [6] Agusta, Yudi, “K-means – penerapan, permasalahan dan metode Terkait”, Jurnal Sistem dan Informatika Vol.3 (Februari 2007): 47-60, 2007.
- [7] Santoso, S. “Statistik Multivariant”, Jakarta: Elex Media Komputindo, 2010.
- [8] Sanoso, B, ”Data Mining: Teknik Pemanfaatan Data untuk Keperluan Bisnis”, Yogyakarta, Graha Ilmu, 2007
- [9] Turban, E., dkk, “Decision Support System dan Intelligent System”, Andi Offset, 2005.
- [10] Witten, Ian H. dan Frank, Eibe, “Data Mining Practical Machine