



IMPLEMENTASI RAPIDMINER PADA ANALISA KETEPATAN BIDANG PEKERJAAN ALUMNI PRODI TEKNIK INFORMATIKA STMIK NURDIN HAMZAH JAMBI

Reny Wahyuning Astuti¹⁾, Sukma Puspitorini²⁾, Fajri Kurnia Akbar³⁾

^{1,2,3} Program Studi Teknik Informatika, STMIK Nurdin Hamza Jambi

e-mail: ¹⁾r3ny4stuti@gmail.com, ²⁾sukm4pit@gmail.com, ³⁾fajrikurnia96@gmail.com

Abstract - Sekolah Tinggi Manajemen Informatika dan Komputer (STMIK) Nurdin Hamzah Jambi is a college that organizes academic education in the discipline of information technology. Every year STMIK always produces scholars who have qualified abilities. However, with the graduation rate in Jambi always increasing, the worries of students who are still running lectures get jobs that are suitable for their field of expertise. Therefore, an analysis is needed to see the level of accuracy of the work obtained by the alumni of Teknik Informatika STMIK Nurdin Hamzah Jambi. The purpose of this study was to find out the level of accuracy of the work of alumni of Teknik Informatika STMIK Nurdin Hamzah Jambi using the K-Means method. This analysis is assisted by using the Rapidminer with its input is the alumni identity, current employment data and expertise data after graduating from Teknik Informatika STMIK Nurdin Hamzah. The process in this analysis is carried out with data that has been given the value of each answer of alumni data in the K-Means method to obtain a Cluster is inappropriate, precise and very precise. The results of this study out of 100 alumni data obtained, 48 alumni entered the Cluster incorrectly, 35 precise Clusters and 17 Clusters were very precise.

Keywords : Alumni, Data mining, Prodi Teknik Informatika, K-Means, STMIK Nurdin Hamzah.

I. PENDAHULUAN

1.1. Latar Belakang Penelitian

Sebagai seorang mahasiswa yang akan memasuki dunia pekerjaan setelah lulus, sudah semestinya mahasiswa yang mengambil jurusan tertentu dapat memanfaatkan ilmu yang diambil selama kurang lebih 4 tahun di bangku perkuliahan untuk diterapkan ketika memasuki dunia pekerjaan. Sekarang ini diperlukan adanya keahlian yang dapat membantu suatu pekerjaan yang akan ditekuni seseorang, terlebih lagi suatu keahlian yang benar-benar dapat menunjang suatu perusahaan atau kantor yang bekerja dalam bidang tertentu.

Sekolah Tinggi Manajemen Informatika dan Komputer (STMIK) Nurdin Hamzah Jambi adalah perguruan tinggi yang menyelenggarakan pendidikan akademik dalam disiplin ilmu teknologi informasi. STMIK Nurdin Hamzah Jambi sendiri memiliki dua program studi yang aktif berjalan saat ini, yakni salah satunya adalah program studi Teknik Informatika. Sama seperti perguruan tinggi pada umumnya, STMIK Nurdin Hamzah Jambi juga berupaya untuk terus memperbaiki mutu pelayanannya guna menghasilkan lulusan yang berkualitas dan kompeten di bidang teknologi informasi serta siap bersaing dalam dunia kerja modern. Hal ini sesuai dengan salah satu misi STMIK Nurdin Hamzah Jambi yakni menghasilkan lulusan yang berkompeten dan berjiwa *technopreneur*.

Setiap tahunnya STMIK selalu menghasilkan sarjana-sarjana yang memiliki kemampuan yang mumpuni dengan bidang-bidang keahlian yang berbeda-beda sesuai dengan minat dan talenta masing-masing sarjana. Terlebih lagi dosen-dosen STMIK selalu mengajarkan kepada mahasiswanya bahwa lulusan STMIK merupakan lulusan yang tidak harus dibawah pemerintah yang biasanya Aparatur Sipil Negara (ASN) tetapi juga harus menjadi seseorang yang dapat membuka suatu lapangan pekerjaan sendiri. Dilain hal saat ini tidak semua orang mendapati pekerjaan yang sesuai bidang keahlian mereka, beberapa faktor dapat membuat seseorang untuk mendapati pekerjaan tetap atau tidak tetap atau untuk mendapatkan pekerjaan yang benar-benar jauh dari bidang mereka. Maka dari itu diperlukannya sebuah analisa yang dapat menunjukkan tingkat ketepatan bidang pekerjaan yang didapat alumni Prodi Teknik Informatika STMIK Nurdin Hamzah. Dengan menggunakan aplikasi Data Mining RapidMiner diharapkan dapat digunakan untuk melihat sejauh mana lulusan STMIK Nurdin Hamzah dapat mencapai target yang diharapkan atau tidak.

Berdasarkan latar Belakang tersebut maka penulis tertarik untuk melakukan penelitian yang berjudul : **“IMPLEMENTASI RAPIDMINER PADA ANALISA KETEPATAN BIDANG PEKERJAAN ALUMNI PRODI TEKNIK INFORMATIKA STMIK NURDIN HAMZAH JAMBI”**.

1.2. Perumusan Masalah

Dari masalah yang telah dijelaskan di atas, masalah yang dapat dirumuskan dalam penelitian ini adalah: “Bagaimana mengimplementasikan RapidMiner pada analisa ketepatan bidang pekerjaan alumni Prodi Teknik Informatika STMIK Nurdin Hamzah Jambi?”

1.3. Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah menganalisa hasil pengelompokan data alumniberdasarkan ketepatan bidang pekerjaan alumni prodi Teknik Informatika STMIK Nurdin Hamzah melalui aplikasi RapidMiner.

II. TINJAUAN PUSTAKA

2.1. Data Mining

Data mining (Connolly dan Begg, 2010) adalah suatu proses ekstraksi atau penggalian data yang belum diketahui sebelumnya, namun dapat dipahami dan berguna dari database yang besar serta digunakan untuk membuat suatu keputusan bisnis yang sangat penting. Sedangkan menurut Han, Jiawei (2011:36), data mining adalah proses menemukan pola yang menarik, dan pengetahuan dari data yang berjumlah besar.

2.2. Clustering

Clustering atau klusterisasi adalah metode pengelompokan data. Menurut Tan, 2006 clustering adalah sebuah proses untuk mengelompokkan data kedalam beberapa cluster atau kelompok sehingga data dalam satu cluster memiliki tingkat kemiripan yang maksimum dan data antar cluster memiliki kemiripan yang minimum. Clustering merupakan proses partisi satu set objek data ke dalam himpunan bagian yang disebut dengan cluster. Objek yang didalam cluster memiliki kemiripan karakteristik antar satu sama lainnya dan berbeda dengan cluster yang lain. Clustering sangat berguna dan bisa menemukan group atau kelompok yang tidak dikenal dalam data. Clustering banyak digunakan dalam berbagai aplikasi seperti misalnya pada business intelligence, pengenalan pola citra, web search, bidang ilmu biologi, dan untuk keamanan (security).

2.3. K-Means

K-Means adalah suatu metode penganalisaan data atau metode Data Mining yang melakukan proses pemodelan tanpa supervisi (unsupervised) dan merupakan salah satu metode yang melakukan pengelompokan data dengan

sistem partisi. Metode k-means berusaha mengelompokkan data yang ada ke dalam beberapa kelompok, dimana data dalam satu kelompok mempunyai karakteristik yang sama satu sama lainnya dan mempunyai karakteristik yang berbeda dengan data yang ada di dalam kelompok yang lain. Dengan kata lain, metode ini berusaha untuk meminimalkan variasi antar data yang ada di dalam suatu cluster dan memaksimalkan variasi dengan data yang ada di cluster lainnya.

K-Means Clustering ini secara umum dilakukan dengan algoritma dasar sebagai berikut:

1. Tentukan jumlah cluster
2. Alokasikan data ke dalam cluster secara random
3. Hitung centroid/rata-rata dari data yang ada di masing-masing cluster
4. Alokasikan masing-masing data ke centroid/rata-rata terdekat
5. Kembali ke Step 3, apabila masih ada data yang berpindah cluster atau apabila perubahan nilai centroid, ada yang di atas nilai threshold yang ditentukan atau apabila perubahan nilai pada objective function yang digunakan di atas nilai threshold yang ditentukan.

III. HASIL DAN PEMBAHASAN

3.1. Proses Penyelesaian Klusterisasi K-Means

Pada tahap ini peneliti melakukan perhitungan manual dengan proses awal mencari cluster data dengan metode K-Means.

Tabel 1. Tabel Normalisasi

No	NP1	NP2	NP3	NP4	NP5	NP6	NP7	NP8
1	0,29	0,00	0,06	0,13	0,00	0,50	0,33	0,00
2	0,00	0,22	0,14	0,00	0,00	0,25	0,33	0,00
3	0,00	0,00	0,41	0,00	0,00	1,00	0,33	0,00
4	0,07	0,11	0,40	0,04	0,00	0,75	0,00	0,00
5	1,00	0,33	0,36	0,00	0,17	0,75	0,33	0,63
6	0,36	0,00	1,00	0,91	0,00	0,00	0,00	0,00
7	0,00	0,33	0,00	0,00	0,33	0,50	1,00	0,00
8	1,00	0,11	0,00	0,09	0,00	0,75	0,33	0,00
9	0,07	0,00	0,93	0,00	0,00	1,00	0,33	0,00
10	0,43	0,22	0,10	0,09	0,50	0,50	0,67	0,00
...
100	0,07	0,00	0,20	0,04	0,50	0,38	0,33	0,00



Setelah data didapat dan dinormalisasikan, data akan siap di *cluster* yang dibagi menjadi 3 kelompok yaitu C1 (kurang tepat), C2 (tepat) dan C3 (sangat tepat) dengan dipersiapkan *initial cluster center* terlebih dahulu dengan nilai yang diacak.

Tabel 2. Tabel *Initial Cluster Center*

	PERTANYAAN KUISIONER							
	P1	P2	P3	P4	P5	P6	P7	P8
C3	0,36	0,11	0,93	0,09	0,00	1,00	0,00	0,50
C2	0,36	0,22	0,47	0,09	0,50	0,50	0,67	0,75
C1	1,00	0,11	0,06	0,00	0,33	0,00	0,33	0,63

Setelah *initial cluster center* telah didapat, selanjutnya dilakukan perhitungan jarak antara nilai dengan rumus yang digunakan pada tiap *Cluster* dengan rumus dibawah ini : $\sqrt{(NP-CP1)^2 + (NP2-CP2)^2 + (NP3-CP3)^2 + (NP4-CP4)^2 + (NP5-CP5)^2 + (NP6-CP6)^2 + (NP7-CP7)^2 + (NP8-CP8)^2}$.

Setelah mendapatkan *cluster center* yang baru akan dilakukannya perulangan iterasi hingga mendapatkan *cluster center* yang baru bernilai sama dengan *cluster center* yang sebelumnya :

Tabel 3. *Cluster Center Initial*

	PERTANYAAN KUISIONER							
	P1	P2	P3	P4	P5	P6	P7	P8
C3	0,36	0,11	0,93	0,09	0,00	1,00	0,00	0,50
C2	0,36	0,22	0,47	0,09	0,50	0,50	0,67	0,75
C1	1,00	0,11	0,06	0,00	0,33	0,00	0,33	0,63

Tabel 4. *Cluster Center Iterasi 1*

	PERTANYAAN KUISIONER							
	P1	P2	P3	P4	P5	P6	P7	P8
C3	0,35	0,09	0,36	0,16	0,17	0,61	0,33	0,01
C2	0,25	0,20	0,40	0,08	0,43	0,49	0,37	0,76
C1	0,88	0,19	0,29	0,13	0,49	0,53	0,28	0,71

Tabel 5. *Cluster Center Iterasi 6*

	PERTANYAAN KUISIONER							
	P1	P2	P3	P4	P5	P6	P7	P8
C3	0,35	0,09	0,36	0,16	0,17	0,61	0,33	0,01
C2	0,25	0,20	0,40	0,08	0,43	0,49	0,37	0,76
C1	0,88	0,19	0,29	0,13	0,49	0,53	0,28	0,71

Setelah dilihat *cluster center* iterasi 6 memiliki nilai yang sama dengan *cluster center* 5. Dengan begitu proses pencarian dihentikan.

Tabel 6. *Cluster Data Untuk Naive Bayes*

No	Nama	Custer
1	Rega Sanjani Alfiqri	C3
2	Syafrizal Edy Chandra	C3
3	Abdy Setiawan	C3
4	Supriyanto	C3

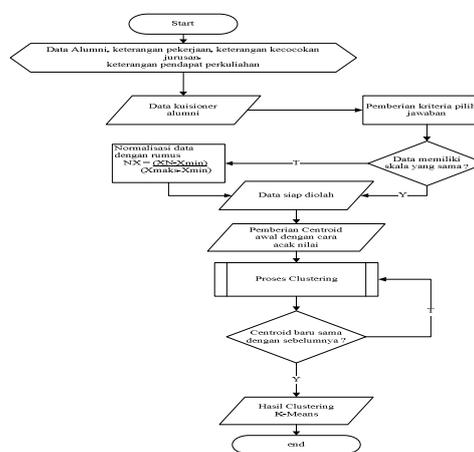
5	Elvi Yanti	C1
6	Panji	C3
7	Angga Pratama	C3
8	M. Jumain	C3
9	Nazratul Putri	C3
10	M. Habib Daiman	C3
...
100	Ezi Mulyadi	C1

3.2. Perancangan Sistem

1. Flowchart

Pada *Flowchart* proses inidimulai dengan proses mempersiapkan data agar dapat diolah dalam metode *K-Means* dengan memberikan nilai tiap atribut. Setelah itu data yang siap diproses dan telah dinormalisasikan akan diproses hingga mendapatkan *Cluster* yang terbagi 3, yaitu *cluster*1 yang termasuk dalam golongan Kurang Tepat, *cluster* 2 yang termasuk dalam Tepat dan *cluster* 3 yaitu yang memiliki ketepatan pekerjaan Sangat Tepat.

Sebelum mendapatkan *cluster – cluster* tersebut diperlukannya sebuah *initial cluster Center* yaitu nilai *initial clustercentre* pada iterasi yang pertama (perhitungan pertama kali) diberikan secara acak. Pada iterasi selanjutnya, nilai *initial clustercentre* (pengulangan ke-1 sampai dengan posisi normal/maksimal iterasi) diberikan dengan menghitung nilai rata-rata data pada setiap *clusternya*. Jika nilai *initial cluster centre* yang baru sama dengan nilai *initial cluster centre* yang lama, maka proses pengelompokan dapat dihentikan. Untuk lebih jelasnya *flowchart* proses ini dapat dilihat dibawah :



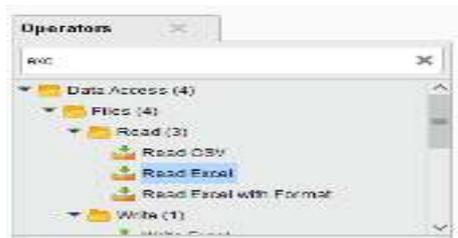
Gambar 1. *Flowchart Proses Clustering*

2. Antarmuka Pada RapidMiner

Tahapan implementasi merupakan tahapan dimana proses penyelesaian dengan *K-Means* diuji dan dioperasikan sesuai dengan data dan hasil yang diinginkan. Berikut ini adalah penjelasan mengenai implementasi RapidMiner pada analisa ketepatan bidang pekerjaan alumni prodi Teknik Informatika STMIK Nurdin Hamzah.

a. Proses Menghubungkan Data

Pada proses ini dilakukan proses penghubung data dari Microsoft Excel dengan aplikasi RapidMiner. Proses dapat dilihat pada gambar dibawah ini :



Gambar 2. Operator Perintah Membaca Excel

Pada gambar diatas terdapat folder *data access* yang memiliki folder *files* dan folder *write* yang memiliki perintah menambah data didalam masing – masing folder tersebut, dalam proses kali ini peneliti menggunakan operator *read excel* yang berfungsi untuk membaca data *excel* untuk dimasukkan kedalam RapidMiner.



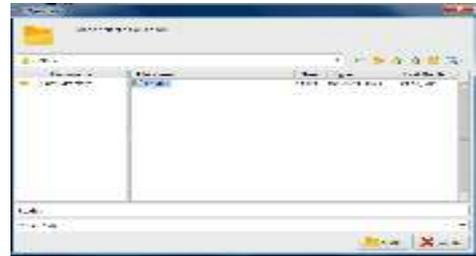
Gambar 3. Process Operator Yang Digunakan

Gambar diatas menunjukkan operator *read excel* berhasil dimasukkan kedalam proses.



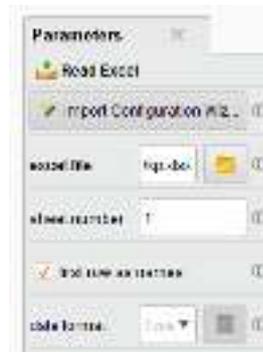
Gambar 4. Parameter Untuk Mengambil Data Excel

Setelah operator *read excel* terdapat dikolom proses untuk memasukkan data *excel* terlebih dahulu kita klik ikon folder yang terdapat diparameter.



Gambar 5. Open File Excel

Setelah mengklik ikon folder di parameter akan muncul jendela *open file* cari data yang akan dipakai.



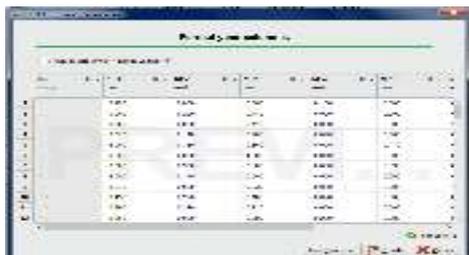
Gambar 6. Import Configuration

Setelah data diambil, klik *import configuration* yang terdapat di parameter untuk melakukan konfigurasi data yang dipakai untuk mengatur format dan pengaturan dari excel diubah kedalam format dan pengaturan RapidMiner.



Gambar 7. Select Cell Yang Akan Di Import

Pada gambar diatas menunjukkan kolom yang terdapat pada data excel peneliti beserta isinya untuk digunakan dalam proses data mining yang dilakukan didalam RapidMiner.



Gambar 8. Format Kolom Data

Setelah memilih data mana saja yang akan digunakan seperti gambar sebelumnya kali ini dilakukan konfigurasi untuk menentukan format data dalam kolom yang akan digunakan. Setelah mengikuti langkah dari gambar diatas data excel yang dimiliki sudah terhubung dengan RapidMiner.

b. Proses Pemilihan Metode

Pada pemilihan metode ini peneliti memilih metode *k-means* dengan mengambil metodenya di RapidMiner pada pemilihan operator. Dapat dilihat pada Gambar 9 berikut ini :



Gambar 9. Metode *K-Means* Yang Dipilih

Setelah data excel dapat digunakan selanjutnya temukan operator *k-means* dengan cara yang sama dengan pengambilan operator *read excel* sebelumnya dan dimasukkan kedalam kolom proses dan dihubungkan dengan operator *read excel*.



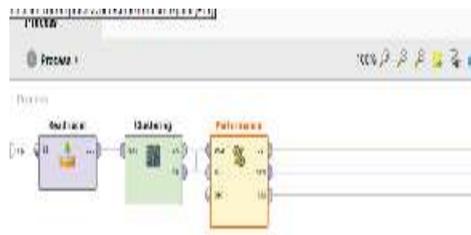
Gambar 10. Parameter Metode *K-Means*

Langkah selanjutnya adalah mengkonfigurasi parameter proses *k-means* untuk menambah *cluster attribute* dan *add as label* yang bertujuan untuk menunjukkan hasil *cluster* yang

akan dihasilkan nanti dengan dipanggil sebagai *label*.

c. Proses Akhir

Pada proses terakhir ini akan ditambah operator *Cluster Distance Performance* sehingga didapatkan jarak antar *cluster*. Gambar dapat dilihat dibawah ini:

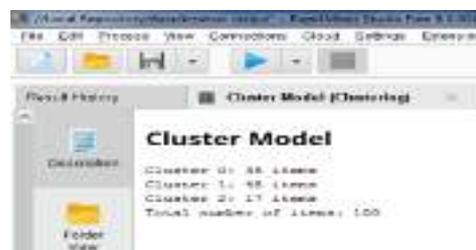


Gambar 11. Proses Akhir

Setelah operator *read excel*, *clustering* dan *performance* dihubungkan seperti gambar diatas proses siap dijalankan.

d. Tampilan Hasil Proses

Pada tampilan hasil proses ini menunjukkan hasil olahan dengan menggunakan metode *K-means*. Adapun antarmukanya dapat dilihat pada gambar berikut ini :



Gambar 12. Hasil Pembagian *Cluster* Melalui *Description*

Pada gambar diatas dapat dilihat bahwa 100 data yang dimiliki terbagi 3 *cluster* yang dalam pembagiannya :

- 35 data masuk kedalam *Cluster_0*
- 48 data masuk kedalam *Cluster_1*
- 17 data masuk kedalam *Cluster_2*

e. Tampilan Tabel Centroid

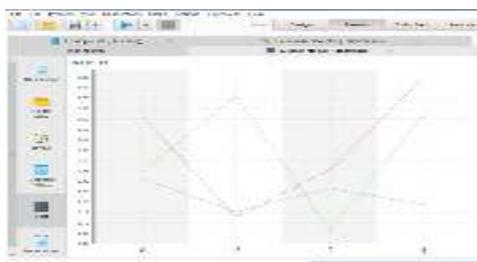
Pada tampilan hasil proses ini menunjukkan *centroid* yang dimiliki tiap *cluster*. Gambar dapat dilihat berikut ini :

Cluster	Centroid X	Centroid Y	Centroid Z
1	1.11	1.22	1.33
2	1.44	1.55	1.66
3	1.77	1.88	1.99
4	2.11	2.22	2.33
5	2.44	2.55	2.66
6	2.77	2.88	2.99

Gambar 13. Tabel Centroid

f. Tampilan Grafik Batang

Pada tampilan hasil proses ini menunjukkan hasil olahan data dengan menggunakan metode *k-means* dalam bentuk grafik batang. Gambar dapat dilihat pada antarmuka berikut ini :



Gambar 14. Grafik Batang Cluster

Pada grafik di atas menunjukkan jarak *cluster* yang dimiliki tiap kolom data yang digunakan.

g. Tampilan Cluster Dan Jarak Cluster

Pada tampilan hasil proses ini menunjukkan hasil olahan dengan menggunakan metode *k-means* dalam bentuk data tiap individu dalam bentuk tabel yang memiliki kelompok *cluster* dan jarak tiap *cluster*. Adapun antarmuka proses dapat dilihat pada gambar dibawah ini :

Gambar 15. Hasil Pembagian Cluster Dan Jarak Antar Cluster

Hasil dari pembagian *cluster* yang dilakukan ditambah kedalam kolom bernama label yang menunjukkan pembagian *cluster* tiap data dan ditunjukkan melalui jarak antara tiap data yang digunakan dalam kolomnya masing – masing.

IV. KESIMPULAN DAN SARAN

4.1. Kesimpulan

Setelah melakukan analisa mendalam metode data mining *K-Means Clustering* yang bertujuan untuk mengelompokkan ketepatan bidang pekerjaan alumni prodi TI STMIK Nurdin Hamzah Jambi, dapat ditarik kesimpulan sebagai berikut :

1. Hasil *clustering* dengan metode *K-Means* pengelompokkan ketepatan bidang pekerjaan alumni dari 100 data alumni didapat 48 alumni masuk kedalam *cluster* kurang tepat, 35 alumni masuk kedalam *cluster* tepat dan 17 alumni masuk kedalam *cluster* sangat tepat.
2. Penggunaan Aplikasi data mining dengan RapidMiner dapat dijadikan sebagai salah satu alternatif untuk mengolah dan menganalisa data yang cukup besar.

4.2. Saran

Dari kesimpulan diatas dan pembahasan pada bab sebelumnya, dapat disarankan beberapa hal, yaitu :

1. Analisa data mining pada RapidMiner dengan metode *K-Means* dapat dibantu dengan pencarian secara manual, dengan maksud untuk memastikan bahwa hasil pencarian baik dengan menggunakan perhitungan manual dan implementasi dengan *software* RapidMiner menunjukkan hasil yang sama.
2. Penelitian ini dapat dikembangkan dengan memperluas area penelitian dan bukan hanya diterapkan untuk menganalisa data alumni prodi Teknik Informatika saja namun juga melibatkan alumni dari prodi lain, seperti Sistem Informasi.
3. Penelitian dapat dikembangkan dengan metode lain untuk melakukan *Clustering* misalnya dengan metode pengklusteran tak terawasi (*unsupervised self organizing map* atau dengan metode *fuzzy subtractive Clustering* untuk memperoleh *Cluster* yang lebih optimal.

DAFTAR REFERENSI

[1] Astuti RW, 2012, *Bahan Ajar Data Mining*, Prodi Teknik Informatika STMIK Nurdin Hamzah, Jambi.

[2] Hermawati, FA, 2009, *Data Mining*, Penerbit Andi, Yogyakarta.



- [3] Jurnal Akademika Vol 11 No. 1, November 2018 p-ISSN : 1970-3984, e-ISSN : 2541-1760, Penentuan Produk Unggulan *Online Shop* Menggunakan *K-Means* dan *Subtractive Clustering*; Astuti RW, Puspitorini, S, Trimastuti D.
- [4] Kusriani, Lutfi ET, 009, *Algoritma Data Mining*, Penerbit Andi, Yogyakarta.
- [5] Nofriansyah D, 2014, *Data Mining VS Sistem Pendukung Keputusan*, Penerbit Deepublish, Yogyakarta.
- [6] Suyanto, Dr, 2017, *Data Mining Untuk Klasifikasi dan Klusterisasi Data*, Penerbit Informatika, Bandung.
- [7] Sulianta F, Juju D, 2010, *Data Mining Meramalkan Bisnis Perusahaan*, Elex Media Komputindo, Jakarta.
- [8] Vuldari, RT, 2017, *Data Mining Teori dan Aplikasi RapidMiner*, Penerbit Gava Media, Yogyakarta.
- [9] http://simki.unpkediri.ac.id/mahasiswa/file_artikel/2017/4d81796c74686bf023266b208d0f553a.pdf (diakses 25 September 2019).
- [10] <https://informatikalogi.com/algoritma-k-means-clustering/> (diakses 25 September 2019).
- [11] <https://informatikalogi.com/algoritma-naive-bayes/> (diakses 25 September 2019).
- [12] <https://docplayer.info/61407944-Penerapan-algoritma-naive-bayes-untuk-memprediksi-profil-lulusan-stmik-lpkia.html> (diakses 25 September 2019).
- [13] <https://snast.akprind.ac.id/ocs236/index.php/snast/2018/paper/download/389/34> (diakses 25 September 2019)

IDENTITAS PENULIS

Nama : Reny Wahyuning Astuti
 NIDN : 1016057803
 TTL : Bajubang, 16 Mei 1978
 Gol/Pangkat : IIIC
 Jab. Fungsional : Lektor
 Email : r3ny4stuti@gmail.com

Nama : Sukma Puspitorini
 NIDN/NIK : 1001048201
 TTL : Blora, 01 April 1982
 Gol/Pangkat : III A
 Jab. Fungsional : Asisten Ahli
 Email : sukma4pit@gmail.com

Nama : Fajri Kurnia Akbar
 NIK/NIDN : -
 Jab Fungsional : -
 Email : fajrikurnia96@gmail.com