

**Pengelompokan Data Kategorik Dengan Algoritma *Robust Clustering Using Links*  
(Studi Kasus: PT. Prudential Life Jalan MT. Haryono Samarinda)**

***Categorical Scale Data Grouping With Robust Clustering Using Links Algorithm  
(Case Study: PT. Prudential Life Road MT. Haryono Samarinda)***

**Isma Dewi<sup>1</sup>, Syaripuddin<sup>2</sup>, dan Memi Nor Hayati<sup>3</sup>**

<sup>1,2</sup>Laboratorium Statistika Komputasi, FMIPA, Universitas Mulawarman

<sup>3</sup>Laboratorium Statistika Terapan, FMIPA, Universitas Mulawarman

<sup>1</sup>E-mail: [ismhadhewi@gmail.com](mailto:ismhadhewi@gmail.com)

**Abstract**

*Cluster analysis is a technique of data mining that is used to group data based on the similarity of attributes of data objects. The problem that is often encountered in cluster analysis is the data on a categorical scale. Categorical scale data grouping can be done using the ROCK (RObust Clustering using linKs) algorithm. The ROCK algorithm is included in the of agglomerative hierarchical clustering algorithms in cluster analysis. This algorithm introduces a concept called neighbors and links in grouping data. Categorical data grouping with ROCK algorithm is done in three steps. The first step is counting similarities. The second step is determining the neighbors and the last is calculating the links between the observation objects. The value of the link is affected by  $\theta$ . The optimum number of clusters in the ROCK algorithm is selected using a minimum ratio value of  $S_w/S_B$ . The purpose of this study is to group 100 data of insurance customers of PT. Prudential Life Samarinda in 2018. Based on the analysis results, obtained that the optimum group is at  $\theta = 0.1$  with a ratio value of  $S_w/S_B$  is 0.1371. The optimum number of groups formed is 2 clusters. The first group consisted of 42 customers and the second group consisted of 58 customers.*

**Keywords:** ROCK Algorithm, Cluster Analysis, Categorical Data

**Pendahuluan**

Analisis kelompok (*cluster analysis*) merupakan salah satu metode yang ada dalam analisis multivariat. Tujuan dari analisis ini adalah mengelompokkan obyek-obyek pengamatan menjadi beberapa kelompok berdasarkan variabel-variabel yang diamati, sedemikian hingga obyek-obyek yang terdapat pada kelompok yang sama mempunyai karakteristik yang relatif homogen dan obyek-obyek antar kelompok mempunyai karakteristik yang berbeda (Sharma, 1996).

Pada awalnya, hampir semua pengembangan metode pengelompokan terkonsentrasi pada data numerik. Dengan berbagai macam konsep jarak yang digunakan sebagai ukuran ketidak miripan (*disimilarity*), misalnya jarak *Euclid*, *miinkowski metric* dan *city-block metric*. Ukuran ketidak miripan yang akan digunakan tergantung kebutuhan. Dalam konteks *database*, banyak data yang berjenis kategori. Sehingga pengelompokan dengan menggunakan metode yang terkonsentrasi pada data numerik memberikan hasil yang kurang memuaskan (Guha dkk, 1999).

Banyak ditemukan algoritma-algoritma yang digunakan untuk mengelompokkan data yang berjenis kategori salah satunya adalah *Robust Clustering using linKs* (ROCK). Algoritma ini dikembangkan untuk mengatasi salah satu masalah penting dalam data *mining*, yaitu mengelompokkan data. Algoritma ROCK merupakan algoritma pengelompokan hirarki *agglomerative* yang

mendasarkan gagasan pengelompokannya pada *link* dan *neighbor*. Sepasang kelompok dikatakan *neighbors* jika mempunyai kemiripan sekurangnya berada dalam ambang batas (*threshold*) tertentu. Banyaknya *link* antar pasangan kelompok adalah banyaknya *common neighbors* atau *neighbors* bersama untuk sepasang kelompok tersebut (Guha dkk, 1999).

Memenuhi kebutuhan yang belum pasti di masa yang akan datang tersebut maka sebagian manusia memerlukan asuransi. Dalam Undang-Undang No. 40 Tahun 2014 Tentang peransuransi menyebutkan bahwa asuransi adalah perjanjian antara dua pihak, yaitu perusahaan asuransi dan pemegang polis, yang menjadi dasar bagi penerimaan premi oleh perusahaan asuransi sebagai imbalan untuk memberikan pembayaran yang didasarkan pada meninggalnya tertanggung dengan manfaat yang besarnya telah ditetapkan dan atau didasarkan pada hasil pengelolaan dana. Salah satu jenis asuransi yang banyak digunakan adalah asuransi jiwa. Asuransi jiwa adalah usaha kerjasama dari sejumlah orang yang sepakat memikul kesulitan keuangan bila terjadi musibah terhadap salah satu anggotanya (Fajriani, 2013).

**Data mining**

Data *mining* merupakan proses ekstraksi informasi dari sekumpulan data yang sangat besar melalui penggunaan algoritma dan teknik

penarikan dalam bidang statistik, pembelajaran mesin dan sistem manajemen basis data (Taruna R dan Saroj, 2013). Selain itu, menurut Mujiasih (2011) data *mining* adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam data berukuran besar. Tan dkk (2006) menyebutkan teknik-teknik data *mining* yang digunakan bertugas untuk menemukan pola baru dan bermakna di dalam basis data yang mungkin masih belum diketahui.

### 1. Clustering

*Clustering* atau analisis *cluster* adalah suatu teknik data *mining* untuk mengelompokkan himpunan objek (*dataset*) ke dalam beberapa grup/*cluster* hanya berdasarkan kemiripan karakteristik dari atribut yang dimiliki oleh data objek sedemikian sehingga data objek yang berada di dalam *cluster* yang sama memiliki kemiripan satu sama lain tetapi mereka tidak mirip dengan data objek yang berada dalam *cluster* yang berbeda (Han dkk, 2012).

Metode dalam analisis *cluster* dibagi menjadi dua yaitu, metode hierarki dan non-hierarki. Metode hierarki memulai pengelompokan dengan dua atau lebih objek yang mempunyai kesamaan paling dekat. Kemudian proses diteruskan ke objek lain yang mempunyai kedekatan dua tipe dasar yaitu *agglomerative* (pemusatan) dan metode *divisive*. Berbeda dengan metode hierarki, metode non-hierarki dimulai dengan terlebih dahulu menentukan jumlah *cluster* yang diinginkan. Setelah jumlah *cluster* diketahui, kemudian proses *cluster* dilakukan tanpa mengikuti proses hierarki (Prasetyo, 2014).

Terkait dengan pengertian dan tujuan dilakukannya analisis *cluster*, dapat dinyatakan bahwa suatu kelompok yang baik adalah kelompok yang memiliki ciri-ciri sebagai berikut (Hair dkk, 2014):

- Homogenitas (kesamaan) yang tinggi antar anggota dalam satu kelompok (*within cluster*).
- Heterogenitas (perbedaan) yang tinggi antar kelompok yang satu tahap dengan kelompok yang lain (*between cluster*).

### 2. Algoritma Robust Clustering using link (ROCK)

Analisis *cluster* untuk data kategori dilakukan menggunakan ukuran kemiripan atau jarak untuk data kategori. Kemudian dapat dilakukan pengelompokan menggunakan metode hierarki maupun non hierarki. Namun, metode hierarki dan non hierarki tradisional dinilai tidak cocok digunakan pada data kategori. Oleh karena itu, telah dikembangkan beberapa metode untuk pengelompokan data kategorik salah satunya adalah algoritma *Robust Clustering using link* (ROCK) yang digunakan dalam penelitian ini (Guha dkk, 1999).

Algoritma ROCK pertama kali diperkenalkan oleh Guha dkk pada tahun 1999. Algoritma ROCK merupakan algoritma *agglomerative hierarchical clustering* untuk data kategori. Algoritma pengelompokan hierarki tradisional yang menggunakan jarak antar titik untuk pengelompokan dinilai kurang sesuai untuk data kategori. Oleh karena itu, dibentuk sebuah konsep baru, yaitu *link* untuk mengukur kesamaan/kedekatan antar sepasang titik data.

Pengelompokan data kategori dengan algoritma ROCK dilakukan dengan tiga langkah sebagai berikut:

- Menghitung *similarity* yaitu

*Similarity* ukuran kemiripan antara pasangan objek ke-*i* dan objek ke-*j* dengan rumusan yang didefinisikan sebagai berikut:

$$\text{sim}(x_i, x_j) = \frac{|Y_i \cap Y_j|}{|Y_i \cup Y_j|}, i \neq j \quad (1)$$

dengan,

$i = 1, 2, 3, \dots, n$  dan  $j = 1, 2, 3, \dots, n$

$Y_i$  = Himpunan pengamatan ke-*i*, dengan

$Y_i = \{x_{i1}, x_{i2}, \dots, x_{mi}\}$ ,

$Y_j$  = Himpunan pengamatan ke-*j*, dengan

$Y_j = \{x_{j1}, x_{j2}, \dots, x_{mj}\}$ ,

$|Y|$  = Bilangan kardinal atau jumlah anggota dari himpunan *Y*.

- Menentukan tetangga

Pengamatan dinyatakan sebagai tetangga jika nilai  $\text{sim}(x_i, x_j) \geq \theta$ .

- Menghitung *link* antar objek pengamatan.

Besarnya *link* dipengaruhi oleh nilai *threshold* (*q*) yang merupakan parameter yang ditentukan oleh peneliti yang dapat digunakan untuk mengontrol seberapa dekat hubungan antar objek. Besarnya nilai *q* yang digunakan adalah  $0 < q < 1$ . Algoritma ROCK berhenti ketika jumlah dari kelompok yang diharapkan sudah terpenuhi atau tidak ada lagi *link* antara kelompok-kelompok. Cara menghitung jumlah *link* untuk semua kemungkinan pasangan dari *n* objek adalah menggunakan matriks **A**. Matriks **A** merupakan matriks berukuran  $n \times n$  yang bernilai 1, jika  $x_i$  dan  $x_j$  dinyatakan mirip (tetangga) dan bernilai 0, jika  $x_i$  dan  $x_j$  dinyatakan tidak mirip (bukan tetangga). Jumlah *link* antar pasangan  $x_i$  dan  $x_j$  diperoleh dari hasil kali antara baris ke  $x_i$  dan kolom ke  $x_j$  dari matriks **A**. Jika *link* antara  $x_i$  dan  $x_j$  semakin besar, maka semakin besar kemungkinan  $x_i$  dan  $x_j$  berada dalam satu kelompok yang sama (Dutta dkk, 2005).

- Penggabungan kelompok

Penggabungan kelompok menggunakan algoritma ROCK didasarkan atas ukuran kebaikan (*goodness measure*) antar kelompok dengan rumus seperti pada Persamaan (7). *Goodness measure* adalah persamaan untuk menghitung jumlah *link* dibagi dengan kemungkinan *link* yang terbentuk berdasarkan ukuran kelompoknya.

$$g(C_i, C_j) = \frac{\text{link}[C_i, C_j]}{(n_i + n_j)^{1+2f(\theta)} - n_i^{1+2f(\theta)} - n_j^{1+2f(\theta)}} \quad (2)$$

Dengan  $\text{link}[C_i, C_j] = \sum_{x_i \in C_i, x_j \in C_j} \text{link}(x_i, x_j)$  yang menyatakan bahwa jumlah *link* dari semua kemungkinan pasangan objek yang ada dalam  $C_i$  dan  $C_j$  dimana (Tyagi & Sharma, 2012):  
 $n_i$  = jumlah anggota dalam kelompok ke- $i$   
 $n_j$  = jumlah anggota dalam kelompok ke- $j$ , dan

$$f(q) = \frac{1-q}{1+q}$$

### 3. Validitas Hasil Clustering

Pengukuran validitas hasil pengelompokan merupakan langkah untuk mengetahui validitas suatu pengelompokan. Kelompok yang baik adalah kelompok yang memiliki kehomogenan yang tinggi antar anggota kelompok dan keheterogenan yang tinggi antar kelompok (Hair dkk, 2014). *Sum of Square Total (SST)* untuk sebuah variabel dengan data kategorik dapat dirumuskan seperti Persamaan (3). Untuk total *Sum of Square Within (SSW)* dirumuskan dalam Persamaan (4). Serta *Sum of Square Between (SSB)* dapat dirumuskan seperti Persamaan (5) (Dewi, 2012).

$$SST = \frac{n}{2} - \frac{1}{2n} \sum_{k=1}^K n_k^2 \quad (3)$$

$$SSW = \frac{n}{2} - \frac{1}{2n} \sum_{c=1}^C \frac{1}{n_c} \sum_{k=1}^K n_{kc}^2 \quad (4)$$

$$SSB = \frac{1}{2} \left( \sum_{c=1}^C \frac{1}{n_c} \sum_{k=1}^K n_{kc}^2 \right) - \frac{1}{2n} \sum_{k=1}^K n_k^2 \quad (5)$$

Sedangkan *Mean of Squares Total (MST)*, *Mean of Squares Within (MSW)*, dan *Mean of Squares Between (MSB)* dapat dirumuskan sebagai berikut:

$$MST = \frac{SST}{(n-1)} \quad (6)$$

$$MSW = \frac{SSW}{(n-C)} \quad (7)$$

$$MSB = \frac{SSB}{(C-1)} \quad (8)$$

$S_w$  (simpangan baku dalam kelompok) dan  $S_B$  (simpangan baku antar kelompok) untuk data kategori dapat dirumuskan sebagai berikut

$$S_w = [MSW]^{\frac{1}{2}} \quad (9)$$

$$S_B = [MSB]^{\frac{1}{2}} \quad (10)$$

Kinerja suatu metode pengelompokan untuk data kategori semakin baik apa bila nilai rasio  $S_w$  dan  $S_B$  semakin kecil. Artinya terdapat homogenitas maksimum dalam kelompok dan heterogenitas maksimum antar kelompok (Bunkers & James, 1996).

$$\frac{S_w}{S_B} \quad (11)$$

### 4. Asuransi

Asuransi merupakan sarana keuangan dalam tata kehidupan rumah tangga, baik dalam menghadapi resiko diri atau benda yang dimiliki. Dalam Undang-Undang No. 40 Tahun 2014 Tentang peransuransi menyebutkan bahwa asuransi adalah perjanjian antara dua pihak, yaitu perusahaan asuransi dan pemegang polis, yang menjadi dasar bagi penerimaan premi oleh perusahaan asuransi sebagai imbalan untuk memberikan pembayaran yang didasarkan pada meninggalnya tertanggung dengan manfaat yang besarnya telah ditetapkan dan atau didasarkan pada hasil pengelolaan dana.

Asuransi berasal dari kata *assurance* atau *insurance*, yang berarti jaminan atau pertanggungan. Hidup penuh dengan ketidakpastian dan manusia selalu berusaha memperkecil atau meminimumkan ketidakpastian tersebut. Misalnya pendidikan yang seseorang tempuh saat ini merupakan salah satu usaha untuk memperkecil ketidakpastian tersebut, dalam usaha mendapatkan jaminan atas perkerjaan yang makin luas (Sembiring, 1986).

### 5. Asuransi Jiwa

Asuransi jiwa adalah usaha kerjasama dari sejumlah orang yang sepakat memikul kesulitan keuangan bila terjadi musibah terhadap salah satu anggotanya. Usaha kerjasama ini dilakukan melalui perusahaan asuransi. Perusahaan yang besar dengan pemegang saham yang banyak akan mudah mengatasi santunan asuransi dari anggota yang meninggal. Setiap orang yang mengansuransikan jiwanya pada suatu perusahaan asuransi berarti sepakat terhadap suatu kontrak tertulis antara dia dan perusahaan. Kontrak tersebut sering disebut polis asuransi (Fajriani, 2013).

### Hasil Penelitian dan Pembahasan

#### 1. Data Penelitian

Data yang digunakan dalam penelitian ini merupakan data sekunder yang diambil dari PT.

Prudential Life Cabang Samarinda berdasarkan dari penelitian Rahmawati, dkk (2019). Adapun variabel yang digunakan dalam penelitian ini yaitu enam variabel bertipe kategorik: Pembayaran premi ( $X_1$ ) memiliki dua kategori yaitu tidak lancar dan lancar. Jenis kelamin ( $X_2$ ) memiliki dua kategori yaitu laki-laki dan perempuan. Status perkawinan ( $X_3$ ) memiliki dua kategori yaitu belum kawin dan kawin. Pekerjaan ( $X_4$ ) memiliki tiga kategori yaitu pegawai swasta, PNS, dan wiraswasta. Penghasilan per bulan ( $X_5$ ) memiliki tiga kategori yaitu 5-10 juta, 11-16 juta, dan >16 juta. Masa Pembayaran ( $X_6$ ) memiliki dua kategori yaitu 1-5 tahun dan 6-10 tahun.

**2. Analisis Cluster Data Kategorik dengan Algoritma ROCK**

Pada data kategorik yang terdiri dari variabel pembayaran premi ( $X_1$ ), jenis kelamin ( $X_2$ ), status perkawinan ( $X_3$ ), pekerjaan ( $X_4$ ), penghasilan per bulan ( $X_5$ ) dan masa pembayaran ( $X_6$ ), nasabah dari PT. Prudential Life Cabang Samarinda tahun 2018 dikelompokkan menggunakan algoritma ROCK. Pada penelitian ini digunakan nilai  $\theta$  sebesar 0,1; 0,2; 0,3; 0,4; 0,5; 0,6; 0,7; 0,8 dan 0,9. Analisis cluster data kategorik menggunakan algoritma ROCK dilakukan dengan bantuan software R versi 3.5.1. adapun tahap-tahapan dalam proses algoritma ROCK untuk 100 cluster adalah sebagai berikut:

1. Inisialisasi data.
2. Menghitung *similarity* antara objek. Contoh perhitungan *similarity* antara objek 1 dan objek 2 adalah sebagai berikut:

$$Y_1 = \{L, Kawin, PNS, 11-16 \text{ Juta, 6-10 Tahun, Lancar}\}$$

$$Y_2 = \{L, Kawin, Pegawai Swasta, 5-10 \text{ Juta, 6-10 Tahun, Lancar}\}$$

$$sim(x_1, x_2) = \frac{|Y_1 \cap Y_2|}{|Y_1 \cup Y_2|} = \frac{4}{8} = 0,5$$

Perhitungan *similarity* dari setiap pasangan objek membentuk matriks **sim** yang merupakan matriks simetris dengan ukuran 100' 100 (banyaknya objek).

3. Menentukan tetangga. Suatu pasangan objek dinyatakan sebagai tetangga jika nilai dari  $sim(x_i, x_j) \geq q$ .

Informasi mengenai hubungan tetangga antar objek pengamatan dapat dinyatakan dengan matriks **A**. Matriks **A** merupakan matriks simetris berukuran 100' 100 yang bernilai 1 jika objek tersebut memenuhi syarat bertetangga dan bernilai 0 jika objek tersebut tidak memenuhi jarak bertetangga.

4. Menghitung *Link*. Link dari pasangan objek ke-i dan objek ke-j dapat dihitung dengan mengalikan baris ke-i dan kolom ke-j dari matriks **A**.
5. Menghitung *Goodness Measure*. Nilai *goodness measure* dapat dihitung menggunakan Persamaan (2). Sebagai contoh

dilakukan perhitungan nilai *goodness measure* untuk objek ke-1 dan objek ke-2 dengan  $q = 0,1$  sebagai berikut:

$$\text{Dengan } f(0,1) = \frac{1 - 0,1}{1 + 0,1} = 0,818$$

$$g(C_1, C_2) = \frac{link[C_1, C_2]}{(n_1 + n_2)^{1+2f(\theta)} - n_1^{1+2f(\theta)} - n_2^{1+2f(\theta)}}$$

$$= \frac{88}{(1+1)^{1+2(0,818)} - 1^{1+2(0,818)} - 1^{1+2(0,818)}}$$

$$= 20,873$$

Pasangan objek atau cluster yang memiliki nilai *goodness measure* paling tinggi akan digabungkan menjadi satu cluster. Dari proses clustering menggunakan algoritma ROCK didapatkan hasil seperti pada Tabel 1.

**Tabel 1.** Validitas Hasil Clustering Data Kategorik

Threshold	Jumlah Cluster yang diinginkan	Jumlah Cluster yang terbentuk	Rasio $S_W / S_B$
0,1	2	1	0
	3	2	0,0437
	4	2	0,5528
	5	3	0,1371
0,2	2	2	$1,4937 \cdot 10^{+15}$
	3	2	2,5654
	4	2	2,5798
0,3	5	3	0,2858
	2	2	$2,5046 \cdot 10^{+15}$
	3	3	2,5785
0,4	4	4	5,7775
	5	3	0,4829
	2	2	$3,3016 \cdot 10^{+15}$
0,5	3	2	2,4437
	4	2	2,3695
	5	2	0,2504
	2	2	$3,1395 \cdot 10^{+15}$
0,6	3	2	1,0529
	4	2	1,0007
	5	2	0,2217
	2	1	0,1898
0,7	3	2	1,0083
	4	3	1,8387
	5	3	0,2037
0,8	2	1	18,9824
	3	2	1,0083
	4	3	1,8387
	5	3	0,2037
0,9	2	17	$4,3588 \cdot 10^{+17}$
	3	17	161,3089
	4	17	161,3089
	5	17	17,8721

**Tabel 1.** Validitas Hasil *Clustering* Data Kategorik (Lanjutan)

<i>Threshold</i>	Jumlah Cluster yang diinginkan	Jumlah Cluster yang terbentuk	Rasio $S_W/S_B$
0,9	2	17	$4,3588 \cdot 10^{+17}$
	3	17	161,3089
	4	17	161,3089
	5	17	17,8721

Berdasarkan Tabel 1 di atas terlihat bahwa kelompok optimum dihasilkan oleh *threshold* 0,1 dengan jumlah kelompok yang terbentuk adalah dua kelompok dan nilai rasio  $S_W/S_B$  sebesar 0,1371.

**Kesimpulan**

Berdasarkan analisis dan pembahasan yang telah dilakukan, maka dapat ditarik kesimpulan sebagai berikut:

1. Hasil pengelompokan data nasabah PT. Prudential Life Cabang Samarinda tahun 2018 dengan menggunakan algoritma ROCK menghasilkan kelompok optimum pada nilai *threshold* 0,1 dan nilai rasio  $S_W/S_B$  sebesar 0,1371. Jumlah *cluster* optimum yang terbentuk adalah sebanyak dua *cluster*. *Cluster* pertama terdiri dari 42 orang nasabah dan kelompok kedua terdiri dari 58 orang nasabah.
2. Karakteristik dari masing-masing *cluster* adalah sebagai berikut:
  - a. Kelompok 1  
Kelompok 1 merupakan kelompok yang terdiri dari 42 orang nasabah. Kelompok ini memiliki karakteristik yang dominan berjenis kelamin laki-laki, berstatus kawin, memiliki pekerjaan PNS dan wiraswasta, berpenghasilan 11-16 juta serta 16 juta keatas, masa pembayaran asuransi sekitar 6-10 tahun dan status pembayaran premi lebih dominan lancar.
  - b. Kelompok 2  
Kelompok 2 merupakan kelompok yang terdiri dari 58 orang nasabah. Kelompok ini memiliki karakteristik yang dominan berjenis kelamin perempuan, berstatus belum kawin, memiliki pekerjaan sebagai pegawai swasta, berpenghasilan 5-10 juta, masa pembayaran asuransi sekitar 1-5 tahun dan status pembayaran premi lebih dominan tidak lancar.

**Daftar Pustaka**

Bunker, M. J., & James, R. M. (1996). Definition of Climate Regions in the Northern Plains Using an Objective Cluster Modification

Technique. *Journal of Climate*, Vol. 9, 130-146.

Dewi, A. (2012). *Metode Cluster Ensemble untuk pengelompokan Desa Perdesaan di Provinsi Riau*. Surabaya: Institut Teknologi Sepuluh November.

Fajriani, A. N., Djuwandi., Wilandari. Y. (2013). Perbandingan Nilai Tebus dan Cadangan Premi pada Asuransi Jiwa Kontinu. Eprints UNDIP.

Guha, S., Rastogi, R., & Shim, K. (1999). ROCK: A Robust Clustering algorithm for Categorical Attributes. *Proceedings of The 15<sup>th</sup> IEEE International Conference on Data Engineering*.

Hair, J. F., Black, W. C., Babin, B. J. & Anderson, R. E. (2014). *Multivariate Data Analysis 7th Edition*. USA: Person Education, Inc.

Han, J., Kamber, M. & Pei, J. (2012). *Data Mining Concepts and Techniques Third Edition*. USA: Elsevier.

Mujiasih, Subekti. (2011). Pemanfaatan Data Mining Untuk Prakiraan Cuaca. *Jurnal Meteorologi dan Geofisika*, Vol. 12, No. 2.

Prasetyo, E. (2014). *Data Mining dan Aplikasi Menggunakan Matlab*. Yogyakarta: ANDI.

Rahmawati, D., Rito, G., & Andi Muhammad Ade, S. (2019). Perbandingan Klasifikasi *Naive Bayes* dan Regresi Logistik pada Data Nasabah Asuransi Tahun 2018, Vol. 4, No. 1.

Sharma, S. (1996). *Applied Multivariate Technique*. New York: John Wiley & Son, Inc.

Tan, P., Steinbach, M. & Kumar, V. (2006). *Introduction to Data Mining*. USA: Person Education, Inc.

Taruna R, Shayara dan Saroj Hiranwal. (2013). Enhanced *Naive Bayes* Algorithm for Intrusion Detection in Data Mining. *International Journal of Computer Science and Information Technologies*, Vol. 4.

