

# ANALISA PREDIKSI EFEK KERUSAKAN GEMPA DARI MAGNITUDO (SKALA RICHTER) DENGAN METODE ALGORITMA ID3 MENGGUNAKAN APLIKASI DATA MINING ORANGE

Lukman Irawan <sup>a,1,\*</sup>, Liyando Hermawan Hasibuan <sup>b,2</sup>, Fauzi <sup>c,3</sup>

<sup>a)</sup> Program Studi Magister Ilmu Komputer, Fakultas Teknologi Informasi, Universitas Budi Luhur  
Jl. Ciledug Raya, Petukangan Utara, Jakarta Selatan 12260

<sup>b)</sup> Program Studi Magister Ilmu Komputer, Fakultas Teknologi Informasi, Universitas Budi Luhur  
Jl. Ciledug Raya, Petukangan Utara, Jakarta Selatan 12260

<sup>c)</sup> Program Studi Magister Ilmu Komputer, Fakultas Teknologi Informasi, Universitas Budi Luhur  
Jl. Ciledug Raya, Petukangan Utara, Jakarta Selatan 12260

<sup>1</sup> lukman.irawan26@gmail.com, <sup>2</sup> liyandohermawan@gmail.com; <sup>3</sup> fauzi.said25@gmail.com

\* corresponding author

## ARTICLE INFO

### Keywords

Klasifikasi  
ID3  
Data mining  
Naives Baiyes  
Gempa bumi

## ABSTRACT

Earthquakes are relatively common natural disasters in Indonesia, mainly due to the interaction of tectonic plates. In this study the seismic energy recorded on the seismograph was measured on the Richter Scale (SR). The dataset collected during Semester 1 of 2019 from the Meteorology and Geophysics, Climatology Agency (BMKG) noted that there were many seismic energy vibrations occurring from small to large around Indonesia. Furthermore, for the attributes of the dataset that have been collected include Date / Time, Longitude, Latitude, Depth (Km), Magnitude (SR), Range of depth (Km) and the effect of earthquake damage selected as the class of the dataset collected, in this study the authors used the method classification with ID3 algorithm to produce effective prediction data for the benefit of earthquake early warning in the Indonesian archipelago.

## 1. Pendahuluan

Gempa bumi merupakan bencana alam yang relatif sering terjadi di Indonesia, terutama akibat interaksi lempeng tektonik. Indonesia merupakan negara kepulauan yang terletak pada pertemuan 4 (empat) lempeng tektonik dunia, yaitu lempeng Eurasia; lempeng Australia; lempeng Pasifik; dan lempeng Filipina. Lempeng Australia dan lempeng Pasifik merupakan jenis lempeng samudera yang bersifat lentur, sedangkan lempeng Eurasia berjenis lempeng benua yang bersifat rigid dan kaku. Pertemuan lempeng tektonik tersebut menyebabkan terjadinya penunjaman serta patahan aktif di dasar lautan dan di daratan. Aktifitas zona tumbukan dan patahan-patahan tersebut berpotensi memicu terjadinya gempa bumi. (Krishna S. Pribadi, dkk, pendidikan siaga bencana ITB. 2008).

Kesiapsiagaan merupakan hal yang penting dan harus dibangun pada setiap kelompok di masyarakat. Pengalaman menunjukkan bahwa kehancuran akibat bencana dapat dikurangi secara drastis jika semua orang lebih siap menghadapi bencana. Sekolah adalah pusat pendidikan yang tidak hanya memberikan ilmu pengetahuan namun juga bekal untuk kelangsungan hidup. Anak-anak merupakan peserta ajar yang paling cepat dan tidak hanya mampu memadukan pengetahuan baru ke dalam kehidupan sehari-hari tetapi juga menjadi sumber pengetahuan bagi keluarga dan masyarakat. Oleh karena itu, pencegahan bencana menjadi salah satu fokus di sekolah dengan memberdayakan anak-anak dan remaja untuk memahami tanda-tanda peringatan bencana dan langkah-langkah yang dapat diambil untuk mengurangi resiko dan mencegah bencana. (KEMENDIKNAS, 2009).

Ukuran gempa dapat dinyatakan dalam skala *Richter* (M) atau skala *Modified Mercalli* (MMI). Skala *Richter* mengukur Magnitude gempa berdasarkan amplitudo yang terjadi sehingga lebih objektif.



**Tabel 1.** *Efek gempa berdasarkan Skala Richter*

Skala Richter	Efek Gempa
< 2,0	Tidak terasa, gempa kecil
2,0 – 2,9	Tidak terasa, terekam alat
3,0 – 3,9	Seringkali terasa, tetapi jarang kerusakan
4,0 – 4,9	Dapat diketahui getarannya, kerusakan tidak signifikan
5,0 – 5,9	Menyebabkan kerusakan pada area kecil
6,0 – 6,9	Dapat merusak area hingga jarak sekitar 160 Km
7,0 – 7,9	Menyebabkan kerusakan pada area luas

Sedangkan skala *Modified Mercalli* mengukur Intensitas gempa berdasarkan efeknya terhadap manusia atau bangunan sehingga lebih bersifat subjektif. karena skala *Mercalli* bersifat subjektif, maka untuk menentukan suatu kerusakan yang diakibatkan oleh gempa, pengamatan yang dilakukan oleh beberapa orang akan mempunyai pendapat yang berbeda mengenai tingkat kerusakannya.

Jika dibandingkan antara skala *Richter* dengan skala *Modified Mercalli* diperoleh hubungan :

**Tabel 2.** Hubungan antara *Magnitudo* Dan *Intensitas* Gempa

Magnitudo ( <i>Richter</i> )	Intensitas (MMI)	Pengaruh-pengaruh Tipikal
≤ 2	I – II	Pada umumnya tidak terasa
3	III	Terasa di dalam rumah, tidak ada kerusakan
4	IV – V	Terasa oleh banyak orang, barang-barang bergerak, Tidak adak kerusakan struktural
5	VI – VII	Terjadi beberapa kerusakan struktural, seperti Retak-retak pada dinding
6	VII – VIII	Kerusakan menengah, seperti hancurnya dinding
7	IX – X	Kerusakan besar, seperti runtuhnya bangunan
≥ 8	XI – XII	Rusak total atau hampir hancur total

*Data mining* diartikan merupakan sekumpulan proses yang berguna mengeksplorasi dan mencari nilai berupa informasi juga relasi – relasi kompleks yang selama ini tersimpan dari suatu basis data. Dengan melakukan penggalian pola informasi terhadap data yang berguna sebagai memanipulasi data menjadi sebuah informasi baru serta lebih bermanfaat yang didapatkan melalui cara mengekstraksi juga mengetahui pola - pola yang berharga atau menarik yang didapatkan dari data didalam basis data.

Data mining digunakan untuk pengelolaan data yang besar serta membantu proses penyimpanan data transaksi dan memproses data warehousing agar kiranya bisa mendapatkan informasi yang diperlukan bagi penggunanya. Dengan data mining kiranya dapat mengetahui dampak atau akibat yang ditimbulkan dari gempa bumi yang terjadi di Indoneisa berdasarkan dari data sampel atau data gempa bumi yang sudah terjadi sebelumnya. Salah satu cara pemanfaatan data mining melalui salah satu teknik data mining yaitu klasifikasi, dengan teknik klasifikasi akan mengelompokan akibat yang ditimbulkan oleh gempa bumi berdasarkan dengan atribut–atribut tertentu. Berdasarkan penelitian-penelitian yang sudah dilaksanakan sebelumnya banyak menggunakan teknik data mining untuk mencari informasi baru yang dapat dimanfaatkan dari data sampel yang sudah ada.

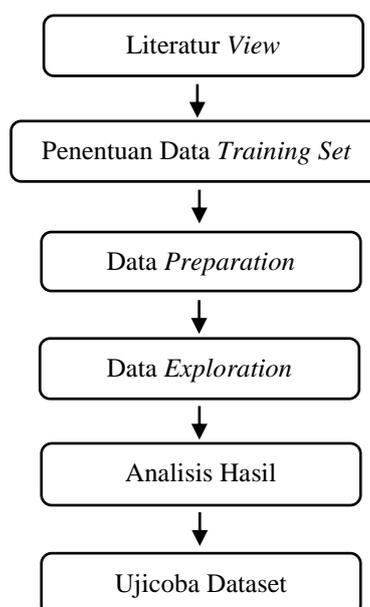
Algoritma *Iterative Dichotomiser 3* (ID3) merupakan salah metode dalam data mining. Data Mining mulai dikenal sejak tahun 1990. ID3 adalah algoritma decision *ID3 learning* (algoritma pembelajaran pohon) yang paling dasar. Algoritma ini melakukan pencarian secara menyeluruh pada semua kemungkinan pohon keputusan. Pembentukan pohon klasifikasi dengan algoritma ID3 melalui dua langkah, yaitu menghitung nilai *entropy* dan menghitung nilai *information gain* dari setiap variabel. ID3 dapat menyelesaikan kasus pada berbagai bidang salah satunya dapat diterapkan pada bidang kesehatan (Santosa, 2007).

*Naïve Bayes Classifier* merupakan sebuah metoda klasifikasi yang berakar pada teorema *Bayes*. Metode pengklasifikasian dg menggunakan metode probabilitas dan statistik yg dikemukakan oleh ilmuwan Inggris *Thomas Bayes*, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya sehingga dikenal sebagai *Teorema Bayes*. Ciri utama dr *Naïve Bayes Classifier* ini adalah asumsi yg sangat kuat (naïf) akan independensi dari masing-masing kondisi / kejadian.

Berdasarkan penelitian terdahulu oleh Efori Buulolo mengatakan “Algoritma data mining dapat memprediksi dampak gempa bumi berdasarkan data seismik yang pernah terjadi dan dampak gempa dipengaruhi oleh beberapa karakteristik atau kondisi gempa yang terjadi” [11]. Selain itu, penelitian lainnya yang telah dilakukan oleh Sandi Fajar Rodiyansyah mengatakan “Algoritma *Naïve Bayes Classifier* memiliki kinerja yang cukup tinggi untuk proses klasifikasi dengan melakukan pengujian sebanyak 13106 terhadap data sampel menghasilkan nilai akurasi terkecil sebesar 78% dan nilai akurasi terbesar 91,60%” [16].

## 2. Metodologi Penelitian

Metodologi penelitian dalam proses klasifikasi atau pengelompokan dengan metode ID3 untuk memprediksi energi seismik yang dipancarkan oleh sumber gempa dimana hal ini meliputi beberapa tahapan, mulai dari literature view, penentuan data training set dan preparation yang terdiri dari data validation, dan transformation dan data reduction serta data exploration. Kemudian, melakukan perhitungan dengan menggunakan aplikasi Orange. Terakhir, penulis melakukan analisis serta pengujian terhadap algoritma yang didapatkan dalam penelitian ini. Berikut ini adalah gambaran flow dan tahapan penelitian yang penulis lakukan untuk penelitian ini:



**Gambar 1.** Flow dan Tahap Penelitian

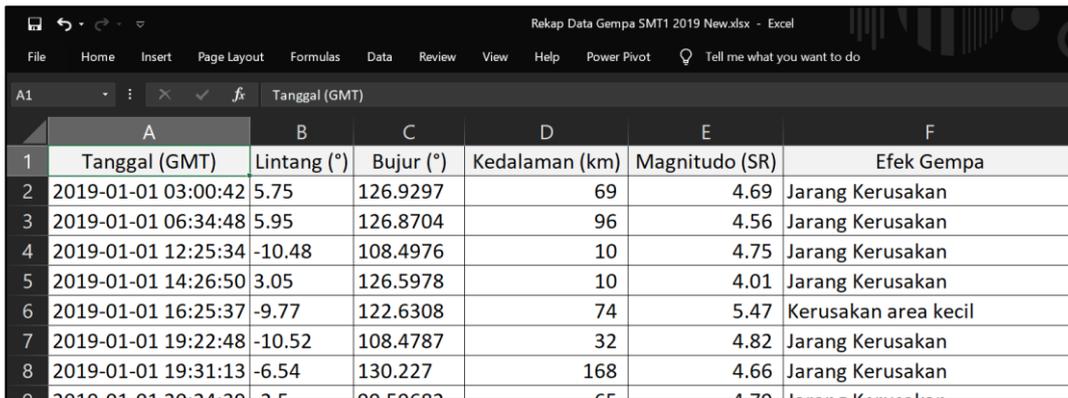
## 2.1. Literatur View

Sebagai literature untuk mendukung penelitian ini, penulis menggunakan beberapa jurnal ilmiah, proceedings, e-book dan website terpercaya. Tulisan-tulisan tersebut diantaranya mengenai *Efek Kerusakan Gempa* seperti, *Earthquake Damage Level of Gorontalo Area Based on Seismicity and Peak Ground Acceleration* [8]. Selain itu penulis juga menggunakan jurnal mengenai *data mining* khususnya klasifikasi seperti *Penerapan Datamining pada Data Gempa Bumi Terhadap Potensi Tsunami di Indonesia* [98], *Optimasi Algoritme Naive Bayes Untuk Klasifikasi Data Gempa Bumi di Indonesia Berdasarkan Hiposentrum* [3]. Dan evaluasi performa algoritma *Klasifikasi Decision ID3 ID3* seperti *Deteksi Gempa Berdasarkan Data Twitter Menggunakan Decision ID3, Random Forest, dan SVM* [15].

## 2.2. Penentuan Data Training Set

Untuk menyelesaikan permasalahan dalam penelitian ini, penulis memulai dengan mengumpulkan data *training set*. Data *training set* tersebut penulis peroleh dari (<http://www.dataonline.bmkg.go.id>) Badan Metereologi, Klimatologi, dan Geofisika (BMKG) dari bulan Januari sampai dengan Juni Tahun 2019. Selanjutnya dari data *training set* tersebut, penulis mulai melakukan pengolahan dengan baik untuk kemudian dilanjutkan ke tahap data preparation.

Jumlah dataset yang penulis dapatkan sebanyak 1.336 *record* data yang berisi informasi diantaranya: Tanggal (GMT), Lintang ( $^{\circ}$ ), Bujur ( $^{\circ}$ ), Kedalaman (Km), Magnitudo (SR), dan Efek Gempa.



	A	B	C	D	E	F
1	Tanggal (GMT)	Lintang ( $^{\circ}$ )	Bujur ( $^{\circ}$ )	Kedalaman (km)	Magnitudo (SR)	Efek Gempa
2	2019-01-01 03:00:42	5.75	126.9297	69	4.69	Jarang Kerusakan
3	2019-01-01 06:34:48	5.95	126.8704	96	4.56	Jarang Kerusakan
4	2019-01-01 12:25:34	-10.48	108.4976	10	4.75	Jarang Kerusakan
5	2019-01-01 14:26:50	3.05	126.5978	10	4.01	Jarang Kerusakan
6	2019-01-01 16:25:37	-9.77	122.6308	74	5.47	Kerusakan area kecil
7	2019-01-01 19:22:48	-10.52	108.4787	32	4.82	Jarang Kerusakan
8	2019-01-01 19:31:13	-6.54	130.227	168	4.66	Jarang Kerusakan

**Gambar 2.** Contoh Tampilan *Dataset* Efek Gempa

Dari atribut tersebut nantinya diharapkan terbentuk sebuah pola yang dapat membantu BMKG untuk membuat prediksi-prediksi kerusakan gempa yang ditimbulkan.

## 2.3. Data Preparation

Tahap selanjutnya adalah penulis melakukan beberapa tahap dalam data *preparation*. Tahap ini meliputi 3 bagian utama yaitu data validation, data transformation, dan data reduction. Sebagai informasi, untuk memudahkan pelaksanaan tahapan ini, penulis menggunakan fitur yang ada di aplikasi Orange. Adapun detail ketiga tahap tersebut adalah:

## 2.4. Data Validation

Penulis melakukan data validation untuk memastikan data training set tersebut dalam kondisi baik dan tidak ada lagi *missing value* di dalamnya. *Missing value* yang dimaksud diantaranya karena ada data yang tidak lengkap, terdapat *outlier* (data abnormal) atau data yang nilainya tidak konsisten. Dari tahapan data validation nantinya akan dihasilkan data yang tidak ada lagi *missing value* nya, dengan kata lain kondisinya normal untuk tahap selanjutnya.

## 2.5. Data Transformation

Tahap berikutnya, penulis melakukan data transformation dengan tujuan untuk menjaga dan memastikan keakuratan data *Training set*. Untuk tahap ini, penulis menerapkan teknik *outlier* yang terdapat dalam aplikasi Orange yaitu fitur *Outlier Detection Method*. Dari proses data transformation ini, nantinya akan dihasilkan data yang lebih akurat untuk di bawa ke tahap selanjutnya.

## 2.6. Data Reduction

Data *reduction* penulis lakukan untuk mengambil sampling data dari *training set* yang ada. Hal ini bertujuan agar set data yang terpilih nantinya, benar-benar siap untuk di klasifikasikan. Dari total 1.336 *record* data.

## 2.7. Data Exploration

Tahap ini merupakan salah satu tahap penting dalam proses *klasifikasi* ini, karena disinilah data yang telah disiapkan sebelumnya akan benar-benar diolah, dihitung nilainya untuk kemudian dianalisis hasilnya. Dalam penelitian ini, penulis melakukan metode pengolahan dan perhitungan data menggunakan fitur di aplikasi Orange. Ada beberapa fitur yang penulis akan gunakan didalamnya seperti fitur outlier dan model *ID3* yang divisualisasikan dengan *ID3 viewer* untuk melihat hasil klasifikasinya.

## 2.8. Analisa Hasil Klasifikasi/Pengelompokan

Pada tahap ini, penulis menganalisis ukuran hasil *kalsifikasi* atau pemisahan yang telah didapatkan dari perhitungan menggunakan aplikasi Orange. Dari tahap ini dapat dijelaskan *atribut* atau *child* terbaik untuk klasifikasi atau pemisahan. Begitu pula dengan pola atau pattern yang mungkin terbentuk dai hasil klasifikasi ini, dapat terlihat dan dianalisis pada tahap ini.

## 2.9. Uji Coba Dataset

Untuk melakukan pengujian dataset, penulis menggunakan metode 10-fold *cross validation* dimana membagi himpunan dataset secara acak menjadi 10 subset yang saling bebas. Selanjutnya dataset *prediction* yang telah penulis peroleh dari Orange dilakukan dihitung dan dianalisis berdasarkan teknik uji coba *ID3* dan *Naïve Bayes*. Penulis melakukan perhitungannya secara manual nilai akurasi dan presisi masing-masing fold kemudian membandingkan kedua hasilnya. Penulis juga mengemukakan *Confusion Matrix* untuk lebih memastikan lagi keakuratan algoritma yang dihasilkan dataset. Idealnya *Confusion matrix* tersebut nilainya sama dengan rata-rata akurasi ataupun presisi *fold* pada 10-fold *cross validation*.

## 3. Hasil dan Pembahasan

Penulis membuat rancangan klasifikasi data mulai dari data training set, tahapan perancangan dan penerapan data mining sehingga nantinya akan dihasilkan suatu pola tertentu yang dapat digunakan untuk memprediksi energi seismik yang dipancarkan oleh sumber gempa. Tahapan tersebut meliputi identifikasi training set data preparation, data exploration baik menggunakan Orange dan analisis serta uji coba hasil perhitungannya.

### 3.1. Data Training Set

Tahap pertama yang penulis laksanakan adalah pengumpulan dan pendefinisian training set data. Adapapun data training set tersebut adalah sebagai berikut:

	Efek Gempa	Kedalaman (km)	Magnitudo (SR)	Lintang (°)	Bujur (°)	Tanggal (GMT)
1	Jarang Kerusakan	69.0	4.69	5.75	126.92969	2019-01-01 03:00:42
2	Jarang Kerusakan	96.0	4.56	5.95	126.87036	2019-01-01 06:34:48
3	Jarang Kerusakan	10.0	4.75	-10.48	108.49755	2019-01-01 12:25:34
4	Jarang Kerusakan	10.0	4.01	3.05	126.59776	2019-01-01 14:26:50
5	Kerusakan area kecil	73.7	5.47	-9.77	122.63078	2019-01-01 16:25:37
6	Jarang Kerusakan	32.4	4.82	-10.52	108.47870	2019-01-01 19:22:48
7	Jarang Kerusakan	168.2	4.66	-6.54	130.22704	2019-01-01 19:31:13
8	Jarang Kerusakan	64.6	4.79	-2.50	99.59682	2019-01-01 20:24:39
9	Jarang Kerusakan	10.0	4.53	6.00	126.91796	2019-01-01 22:07:32
10	Jarang Kerusakan	18.5	4.45	5.84	126.85378	2019-01-01 22:11:18
11	Kerusakan area kecil	41.8	5.20	-1.96	139.10683	2019-01-02 07:59:50
12	Jarang Kerusakan	22.1	4.28	-5.26	130.33926	2019-01-02 08:43:33
13	Jarang Kerusakan	267.3	4.30	-7.23	127.25760	2019-01-02 08:56:25
14	Jarang Kerusakan	124.5	4.55	3.00	128.02165	2019-01-02 12:47:43
15	Jarang Kerusakan	10.0	4.34	-10.38	108.48998	2019-01-02 12:52:05
16	Jarang Kerusakan	24.7	4.25	-6.49	103.63695	2019-01-02 13:33:08
17	Jarang Kerusakan	21.8	4.45	5.56	126.44847	2019-01-02 13:35:59
18	Jarang Kerusakan	21.4	4.62	-6.64	103.46632	2019-01-02 13:40:21
19	Jarang Kerusakan	47.8	4.84	-9.61	118.03931	2019-01-02 14:26:07
20	Jarang Kerusakan	133.4	4.48	5.70	126.00269	2019-01-02 21:26:45
21	Jarang Kerusakan	10.5	4.20	-9.49	118.04832	2019-01-02 23:47:16
22	Jarang Kerusakan	207.2	4.02	-3.99	127.85457	2019-01-03 02:24:19
23	Jarang Kerusakan	13.2	4.75	-3.70	127.82209	2019-01-03 04:21:22
24	Jarang Kerusakan	10.0	4.40	-9.29	109.64494	2019-01-03 07:42:29

Gambar 3. Data File *Training Set* Gempa

Dalam data *Training Set* diatas, yang berisi informasi data gempa untuk periode Januari sampai dengan Juni Tahun 2019, terdapat 1.336 *record* data yang tersedia. *Training Set* tersebut kemudian penulis upload kedalam Orange untuk memudahkan pengolahan data termasuk data *preparation*. Dalam set tersebut, tidak terlihat *Missing Value*, karena kondisi datanya lengkap. Dari tabel data *Trainng Set* diatas pula, dapat diketahui bahwa setidaknya ada 6 atribut yang penulis pilih dari kumpulan data tersebut untuk selanjutnya diolah agar terbentuk suatu pola prediksi. Atribut *Efek Gempa* pada tabel tersebut diketahui sebagai label kelas dengan nilai “*Jarang Kerusakan*”, “*Kerusakan Area Kecil*”, “*Kerusakan Area Luas*” dan “*Kerusakan hingga 160 Km*” yang penulis isi berdasarkan definisi *Efek Gempa* yang telah dibahas diawal. Sementara itu, 4 atribut lainnya adalah :

- Efek Gempa, berisi informasi *record* data Kerusakan gempa yang ditimbulkan dan merupakan atribut kontinu.
- Kedalaman (Km), berisi informasi *record* data ukuran kedalaman gempa dalam satuan Kilometer (Km) dan merupakan atribut kontinu.
- Magnitude (SR), berisi informasi *record* data ukuran gempa dalam satuan *Skala Richter* (SR) dan merupakan atribut kontinu.
- Lintang ( $^{\circ}$ ), berisi informasi *record* data lokasi lintang dan merupakan atribut kontinu.
- Bujur ( $^{\circ}$ ), berisi informasi *record* data lokasi bujur dan merupakan atribut kontinu.
- Tanggal (GMT), berisi informasi *record* data Tanggal dan Jam, kemudian merupakan atribut kontinu.

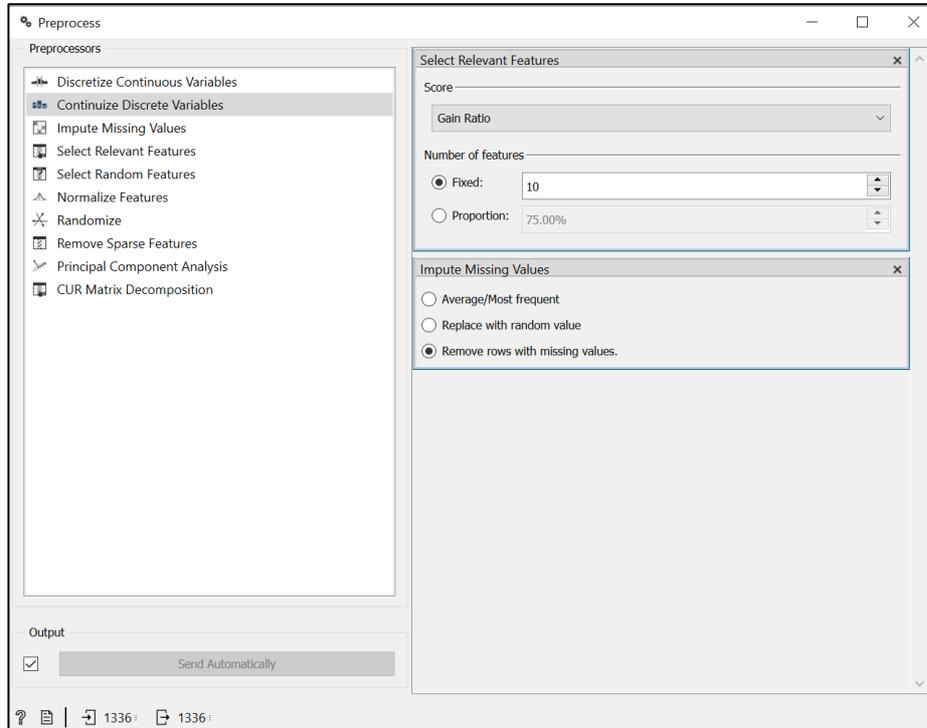
### 3.2. Data Preparation

Bagian ini berisi tentang tahap persiapan data yang terdiri data *validation*, data *transformation* dan data *reduction*. Semua tahap persiapan ini penulis lakukan menggunakan Aplikasi Orange (<https://orange.biolab.si/>) dengan tujuan kemudahan dan keefektifan dalam proses dan pengolahan.

### 3.3. Data Validation

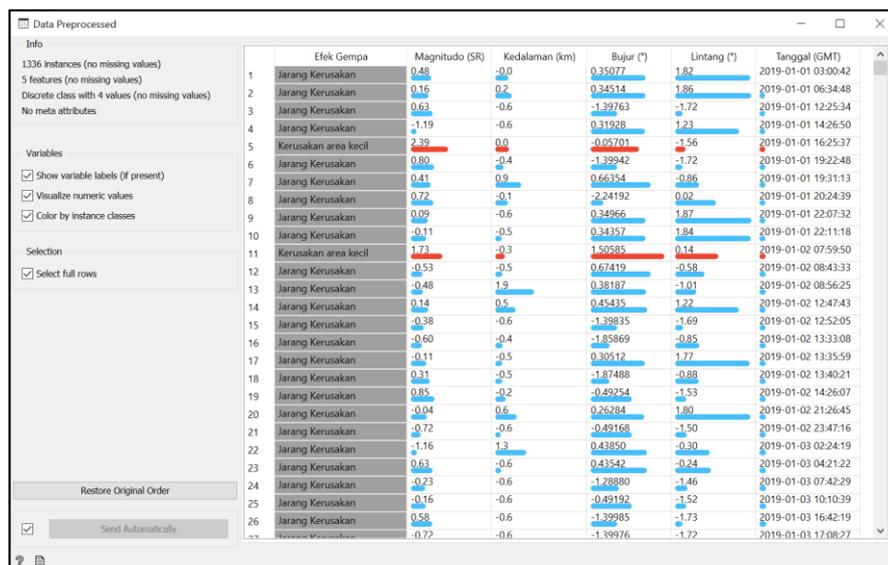
Seperti yang telah dijelaskan sebelumnya bahwa pada dataset yang penulis gunakan sudah baik dengan tidak adanya *Missing Values* berjumlah 1.366 *record* data, untuk meningkatkan validasi data yang ada, selain itu penulis menambahkan fitur Preprocess yaitu *Relevant Feature*, hal ini bertujuan untuk mengetahui *Feature* yang relevant sesuai dengan score *Gini Ratio*. Untuk penelitian ini penulis

tetap menggunakan fitur Preprocess yang terdapat dalam Orange untuk mengkonfirmasi hal-hal tersebut.



**Gambar 4.** Fitur *Preprocess* (*Missing Value and Relevant Feature*)

Selanjutnya ada pula fitur Data Table yang dapat digunakan untuk melihat data hasil validasinya. Hasilnya ditemukan tidak adanya *Missing Value* dalam data *Training Set* sehingga jumlah data tetap 1.336 *record* data dan dapat dilanjutkan ke proses berikutnya.

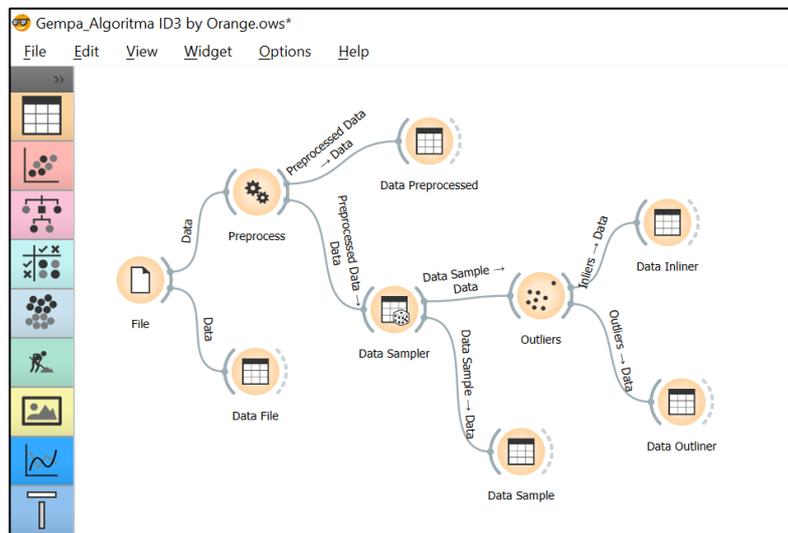


	Efek Gempa	Magnitudo (SR)	Kedalaman (km)	Bujur (°)	Lintang (°)	Tanggal (GMT)
1	Jarang Kerusakan	0.48	-0.0	0.35077	1.82	2019-01-01 03:00:42
2	Jarang Kerusakan	0.16	0.2	0.34514	1.86	2019-01-01 06:34:48
3	Jarang Kerusakan	0.63	-0.6	-1.39763	-1.72	2019-01-01 12:25:34
4	Jarang Kerusakan	-1.19	-0.6	0.31928	1.23	2019-01-01 14:26:50
5	Kerusakan area kecil	2.39	0.0	-0.05701	-1.56	2019-01-01 16:25:37
6	Jarang Kerusakan	0.80	-0.4	-1.39942	-1.72	2019-01-01 19:22:48
7	Jarang Kerusakan	0.41	0.9	0.66354	-0.86	2019-01-01 19:31:13
8	Jarang Kerusakan	0.72	-0.1	-2.24192	0.02	2019-01-01 20:24:39
9	Jarang Kerusakan	0.09	-0.6	0.34966	1.87	2019-01-01 22:07:32
10	Jarang Kerusakan	-0.11	-0.5	0.34357	1.84	2019-01-01 22:11:18
11	Kerusakan area kecil	1.73	-0.3	1.50585	0.14	2019-01-02 07:59:50
12	Jarang Kerusakan	-0.53	-0.5	0.67419	-0.58	2019-01-02 08:43:33
13	Jarang Kerusakan	-0.48	1.9	0.38187	-1.01	2019-01-02 08:56:25
14	Jarang Kerusakan	0.14	0.5	0.45435	1.22	2019-01-02 12:47:43
15	Jarang Kerusakan	-0.38	-0.6	-1.39835	-1.69	2019-01-02 12:52:05
16	Jarang Kerusakan	-0.60	-0.4	-1.85869	-0.85	2019-01-02 13:33:08
17	Jarang Kerusakan	-0.11	-0.5	0.30512	1.77	2019-01-02 13:35:59
18	Jarang Kerusakan	0.31	-0.5	-1.87488	-0.88	2019-01-02 13:40:21
19	Jarang Kerusakan	0.85	-0.2	-0.49254	-1.53	2019-01-02 14:26:07
20	Jarang Kerusakan	-0.04	0.6	0.26284	1.80	2019-01-02 21:26:45
21	Jarang Kerusakan	-0.72	-0.6	-0.49168	-1.50	2019-01-02 23:47:16
22	Jarang Kerusakan	-1.16	1.3	0.43850	-0.30	2019-01-03 02:24:19
23	Jarang Kerusakan	0.63	-0.6	0.43542	-0.24	2019-01-03 04:21:22
24	Jarang Kerusakan	-0.23	-0.6	-1.28880	-1.46	2019-01-03 07:42:29
25	Jarang Kerusakan	-0.16	-0.6	-0.49192	-1.52	2019-01-03 10:10:39
26	Jarang Kerusakan	0.58	-0.6	-1.39985	-1.73	2019-01-03 16:42:19
27	Jarang Kerusakan	-0.72	-0.6	-1.39976	-1.72	2019-01-03 17:08:27

**Gambar 5.** Data *Preprocessed* - *Training Set* Gempa

### 3.4. Data Transformation

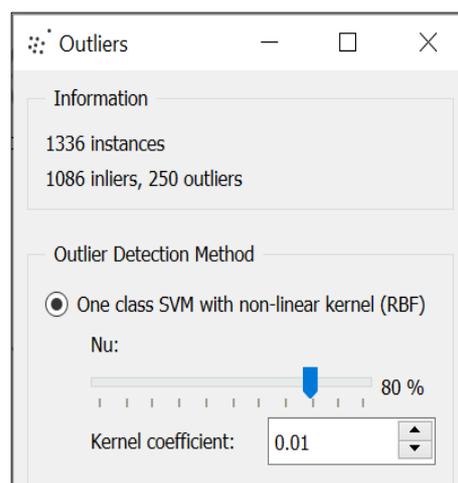
Untuk melakukan proses *Transformasi* data, penulis menggunakan fitur *Outlier* yang terdapat didalam aplikasi Orange. *Oulier* berarti terdapat suatu data yang tidak wajar (anomali) dalam dataset sehingga harus dilakukan perbaikan sebelum ketahap berikutnya. Dibawah ini tampilan flow proses fitur *Outlier* pada aplikasi Orange :



**Gambar 6.** Flow Process *Outlier* didalam aplikasi Orange

Untuk mendeteksi dataset *Outlier* dan ketahap selanjutnya, pada fitur *Outlier* ini penulis menggunakan metode deteksi outlier SVM dengan kernel RBF ini berfungsi untuk menghindari *missclasifikasi* disetiap sample data *training dataset*.

Dari hasil metode SVM dengan kernel RBF ini didapatkan pembagian data *Outlier* sebanyak 205 *record* data dan *Inlier* sebanyak 1.086 *record* data dari total 1.336 *record* data. Dibawah ini tampilan fitur *Outlier*:



**Gambar 7.** *Outlier Detection Methods*

Pada penelitian ini, penulis memilih menggunakan data *Inliner* karena data ini dianggap lebih normal/wajar setelah memisahkan data *Outlier* pada proses *Outlier* sebelumnya.

Selanjutnya, penulis menampilkan data hasil prosesnya dalam Data *Inliers*, dari total data *Training Set* 1.336 *record* data, didapatkan data *Inlier* sebanyak 1.086 *record* data dengan 5 atribut dan 1 kelas. Berikut ini gambaran data *Inlier* sebagai proses *transformasi* dalam aplikasi Orange:

	Efek Gempa	Magnitudo (SR)	Kedalaman (km)	Bujur (°)	Lintang (°)	Tanggal (GMT)
1	Jarang Kerusakan	4.06	39.3	126.90436	3.69	2019-02-27 22:29:17
2	Jarang Kerusakan	4.27	10.0	122.69118	-3.90	2019-01-21 13:54:12
3	Jarang Kerusakan	4.14	10.0	108.49005	-10.49	2019-01-05 22:53:37
4	Jarang Kerusakan	4.68	14.1	138.55588	-2.51	2019-06-26 04:50:00
5	Jarang Kerusakan	4.45	226.8	102.97605	-2.91	2019-01-06 20:19:29
6	Kerusakan area kecil	5.09	10.0	116.49139	-8.44	2019-03-17 07:09:20
7	Jarang Kerusakan	4.25	10.0	117.72352	-8.13	2019-06-12 06:46:57
8	Jarang Kerusakan	4.01	19.4	132.91280	-4.45	2019-03-22 14:15:06
9	Jarang Kerusakan	4.07	102.9	123.49133	-0.04	2019-03-02 01:27:50
10	Jarang Kerusakan	4.86	252.6	122.44542	-0.07	2019-04-16 15:44:25
11	Jarang Kerusakan	4.37	33.8	118.97905	-9.55	2019-01-06 05:37:23
12	Kerusakan area kecil	5.15	82.5	131.61714	-6.31	2019-04-30 02:28:49
13	Jarang Kerusakan	4.26	10.0	122.50877	-2.02	2019-04-12 14:26:49
14	Jarang Kerusakan	4.29	153.5	95.70546	5.43	2019-02-09 09:21:57
15	Jarang Kerusakan	4.26	10.0	119.31281	-9.96	2019-02-12 09:34:51
16	Kerusakan area kecil	5.06	109.8	129.58607	-3.93	2019-02-26 12:15:08
17	Jarang Kerusakan	4.65	46.5	99.11379	-0.48	2019-05-23 03:32:28
18	Jarang Kerusakan	4.88	158.5	125.85466	5.68	2019-05-13 03:59:40
19	Jarang Kerusakan	4.18	16.3	126.99806	1.09	2019-01-23 12:11:08
20	Kerusakan area kecil	5.03	49.8	127.05801	3.78	2019-03-06 13:57:20
21	Jarang Kerusakan	4.36	10.0	122.60268	-2.02	2019-04-12 13:29:39
22	Kerusakan area kecil	5.74	136.2	130.42876	-6.64	2019-01-04 06:22:44
23	Jarang Kerusakan	4.32	69.9	107.85573	-7.77	2019-06-02 04:52:36
24	Jarang Kerusakan	4.99	13.1	138.60912	-2.57	2019-06-24 22:39:47

Gambar 8. Data Sample *Inlier* – Training Set Gempa

Data Sample *Inliers* diatas yang nantinya akan diolah lebih lanjut pada tahap selanjutnya. Sampai pada tahap ini, data benar-benar sudah siap untuk diolah dan dieksplorasi lebih jauh.

### 3.5. Data Exploration

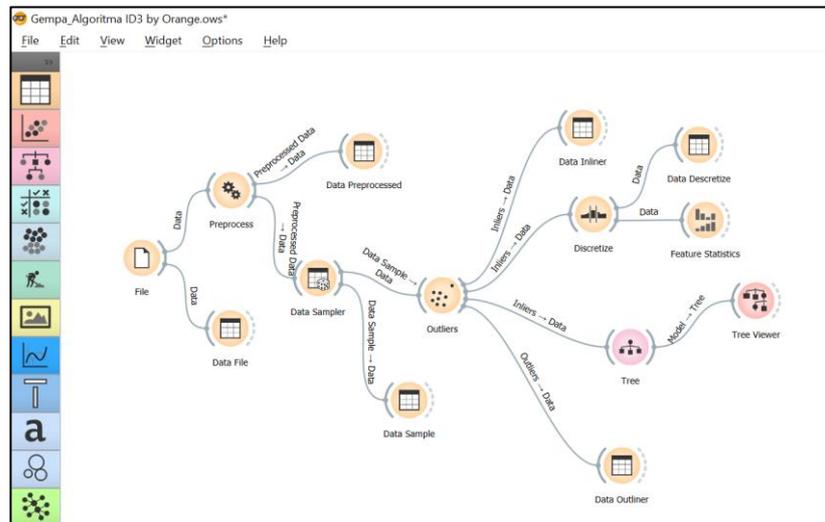
Pada proses pengolahan dan eksplorasi data *Training Set* menggunakan Orange, penulis lebih menonjolkan fitur-fitur yang ada pada Orange dan membuat kalkulasi statistiknya. Selain itu, penulis juga mencoba mengidentifikasi intensitas dari hubungan antar atribut. Pada tahapan ini juga, penulis menguji coba dengan teknik Univariate Analysis. Sebenarnya dalam tahap ini dapat pula dilakukan pengujian dengan teknik Bivariate Analysis dan Multivariate Analysis. Namun karena keterbatasan, penulis hanya menggunakan Univariate Analysis dimana dilakukan investigasi properties dari setiap atribut. Teknik tersebut dilakukan dengan menerapkan fitur “Feature Statistics” yang disediakan oleh Orange. Gambarnya ada sebagai berikut:

Name	Distribution	Center	Dispersion	Min.	Max.	Missing
Efek Gempa	[Histogram]	Jarang Kerusakan	0.37			0 (0%)
Lintang (°)	[Histogram]	-1.85 - 2.29	1.61			0 (0%)
Bujur (°)	[Histogram]	> 129.69935	1.61			0 (0%)
Kedalaman (km)	[Histogram]	< 10.1	1.59			0 (0%)
Magnitudo (SR)	[Histogram]	4.33 - 4.54	1.61			0 (0%)

Gambar 9. Hasil *Feature Statistics* pada Aplikasi Orange

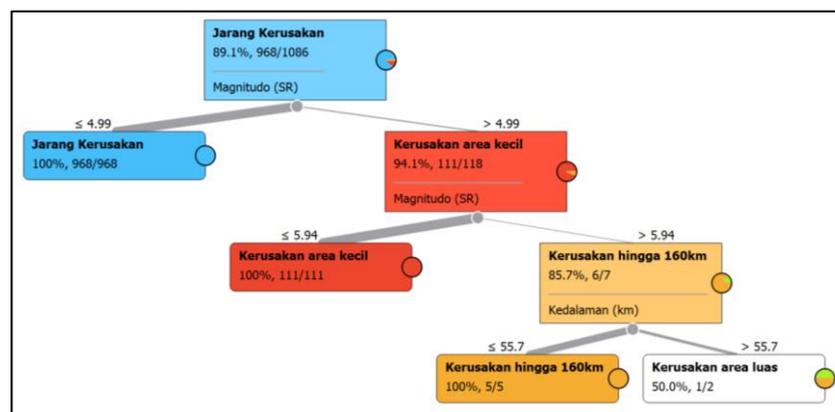
Dari gambar 10 diatas, hasil pengolahan datanya secara umum dapat dibaca. Rata-rata gempa yang ada dalam Data *Training Set* Gempa Efek Gempa *Jarang Kerusakan*, dengan kedalaman < 10 Km, serta *Magnitudo* sebesar 4.33 SR.

Selanjutnya adalah pengolahan data dengan metode klasifikasi dengan tujuan mengelompokkan data film yang *Efek gempa* nya “Jarang Kerusakan”, “Kerusakan area Kecil”, “Kerusakan area Luas”, dan “Kerusakan hingga 160 Km”. Untuk melakukan klasifikasi di Orange, penulis menggunakan fitur “ID3” lalu menampilkan grafik pohon keputusannya dengan fitur “ID3 Viewer” seperti gambar berikut:



**Gambar 10.** Proses Fitur “ID3” dan “ID3 Viewer” Pada Orange

Dengan pengolahan data menggunakan Orange, penulis cukup mudah untuk melihat hasil pohon keputusannya yakni dengan fitur “ID3 Viewer”. Hasil pohon keputusannya kemudian penulis analisis. Untuk lebih jelasnya, penulis cukup mengklik fitur “ID3 Viewer” yang terdapat pada canvas dan potongan hasilnya adalah sebagai berikut:



**Gambar 11.** Grafik Hasil Klasifikasi ID3 Pada Orange

Pada grafik ID3 diatas, hasilnya percabangan yang dihasilkan dalam pohon keputusan juga menjadi banyak, namun tetap dapat dibaca dan dianalisis dengan baik.

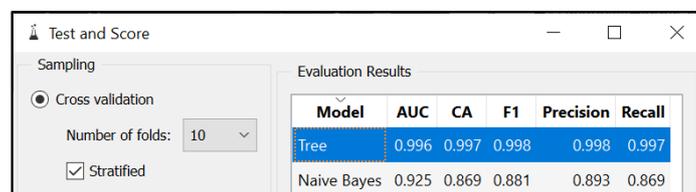
### 3.6. Analisa Hasil Klasifikasi/Pengelompokan

Dari pohon keputusan yang penulis peroleh dalam penelitian ini, mendapatkan pola atau pattern bahwa yang menjadi *root* nya adalah atribut *Efek Gempa*. Jika *Efek gempa* dengan *Magnitudo* < 4,99 SR, maka *Efek gempa* yang dihasilkan adalah *Jarang Kerusakan*. Sedangkan apabila *Efek gempa* dengan *Magnitudo* > 4,99 SR, maka *Efek gempa* yang dihasilkan adalah *Kerusakan area Kecil*, selanjutnya pada *Efek gempa* dengan *Kerusakan area Kecil* apabila *Magnitudo* < 5,94 SR, maka *Efek*

*gempa* yang dihasilkan adalah *Kerusakan area Kecil*, sedangkan jika *Magnitudo* > 5,94 SR, maka *Efek gempa* yang dihasilkan adalah *Kerusakan hingga 160 Km*. Kemudian dari *Efek gempa* dengan *Kerusakan hingga 160 km* apabila  $\leq 55,7$  Km, maka *Efek gempa* yang dihasilkan hanya *Kerusakan hingga 160 Km*, tetapi apabila *Efek gempa* yang dihasilkan *Kerusakan hingga 160 Km* terjadi pada kedalaman > 55,7 Km maka *Efek gempa* yang dihasilkan adalah *Kerusakan area Luas*. Intinya, Orange memilih *Efek gempa* sebagai atribut terbaik untuk dijadikan *root* dan kedalaman ID3 sampai 7 nodes dan 4 leaves.

### 3.7. Uji Coba Dataset

Uji coba dilakukan dengan metode 10-fold *cross validation* dimana membagi himpunan dataset secara acak menjadi 10 *subset* yang saling bebas. Untuk lebih meyakinkan proses uji coba, pada tahapan ini penulis menggunakan teknik uji coba *ID3* dan uji coba *naive bayes*. Untuk lebih memudahkan, penulis menggunakan fitur “*Test & Score*” serta “*Confusion Matrix*” yang ada dalam aplikasi Orange. Pada fitur “*Test & Score*” penulis memilih *cross validation* pada sampling page seperti gambar berikut:



**Gambar 12.** Fitur 10-fold *Cross Validation* Pada Orange

Dari data diatas, terlihat bahwa uji coba teknik *ID3* memiliki hasil pengujian yang lebih baik daripada teknik *Naives Baiyes* terutama pada Akurasi dan score F1. Pada uji coba teknik *Naives Baiyes*, akurasi yang diperoleh sebesar 0,869 (86,9%), presisi sebesar 0.893 (89,3%), recall sebesar 0,997 (99,7%) serta F1 sebesar 0,998 (99,8%). Angka sedikit lebih baik terlihat pada teknik uji coba *ID3* dimana akurasinya sebesar 0,997 (99,7%), presisi sebesar 0,998 (99,8%), recall sebesar 0,997 (99,7) dan F1 sebesar 0,998 (99,8%). Hal ini karena dataset yang penulis gunakan pada penelitian ini banyak dan dan secara teori memang lebih cocok menggunakan uji coba teknik *ID3*.

Meski demikian, berdasarkan hasil pengujian pada gambar 7 diatas, sebenarnya perbedaan kedua teknik ujicoba baik *ID3* maupun *naive bayes* hampir sama. Selain itu, baik akurasi, presisi, recall dan F1 semuanya menunjukkan hasil baik yang artinya bahwa algoritma klasifikasi juga telah berjalan dengan baik. Namun jika melihat Area Under the ROC Curve (AUC) yang menentukan kinerja model algoritma, terlihat bahwa teknik *naive bayes* jauh lebih baik. Model yang baik adalah model yang nilai AUC nya mendekati 1.0 [23]. Sementara, pada gambar hasil pengujian menunjukkan nilai AUC teknik *ID3* 0,996 (99,6%) lebih baik dibanding nilai *Naives Baiyes* yang hanya diangka 0,925 (92,5%). Oleh karena itu, penulis lebih memilih teknik uji coba *ID3* dalam penelitian ini.

Untuk menunjukkan hasil pengujiannya secara umum, penulis kemudian memanfaatkan fitur “*Confusion Matrix*”, hasilnya baik pada ujicoba teknik *ID3* adalah sebagai berikut:

		Predicted				Σ
		Jarang Kerusakan	Kerusakan area kecil	Kerusakan area luas	Kerusakan hingga 160km	
Actual	Jarang Kerusakan	968	0	0	0	968
	Kerusakan area kecil	0	110	1	0	111
	Kerusakan area luas	0	0	0	1	1
	Kerusakan hingga 160km	0	0	1	5	6
Σ		968	110	2	6	1086

**Gambar 13.** *Confusion Matrix* Uji Coba Teknik *ID3*

Pada dasarnya, tabel *Confusion Matrix* diatas merepresentasikan hasil pengujian yang telah dibahas sebelumnya. Hasil pengujiannya menunjukkan bahwa algoritma sudah baik. Dengan

demikian, klasifikasi ID3 yang dilakukan untuk memprediksi *Getaran Gempa/Magnitudo (SR)* yang kemungkinan akan memiliki prediksi *Efek gempa* lebih baik dan dapat digunakan oleh BMKG sebagai salah satu acuan pengambilan keputusan peringatan dini sebelum terjadi kerusakan dari gempa yang dihasilkan.

#### 4. Kesimpulan

Beberapa kesimpulan yang penulis dapat ambil dari penelitian ini antara lain:

- a. Berdasarkan sampel training set data yang ada pada penelitian ini, atribut yang menjadi root dan node selanjutnya berturut-turut adalah Efek Gempa, Magnitudo (SR), Kedalaman (Km), Bujur(0), Lintang (0), dan Tanggal (GMT). Hasil tersebut menjadi pattern yang dapat digunakan BMKG dalam memprediksi kemungkinan tinggi atau rendahnya efek gempa yang dihasilkan.
- b. Berdasarkan algoritma ID3 yang dihasilkan dalam penelitian ini, kemungkinan gempa yang dihasilkan, melalui pengukuran Maginitudo (Skala Richter) akan memberikan informasi Efek kerusakan diantaranya :
  - Jika Magnitudo  $< 4,99$  SR, maka Efek kerusakan yang dihasilkan adalah Jarang Kerusakan.
  - Jika Magnitudo  $> 4,99$  SR, maka Efek kerusakan yang dihasilkan adalah Kerusakan area Kecil.
  - Jika Magnitudo  $> 5,94$  SR, maka Efek kerusakan yang dihasilkan adalah Kerusakan hingga 160 Km dengan kedalaman  $< 55,7$  Km, sedangkan apabila kedalamannya  $> 55,7$  Km maka akan terjadi Efek kerusakan area Luas.
- c. Teknik uji coba ID3 adalah teknik yang menghasilkan nilai paling akurat untuk menguji algoritma yang dihasilkan oleh dataset.
- d. Berdasarkan teknik uji coba ID3 yang dilakukan, didapatkan nilai akurasi pengujian algoritma sebesar 0,997 (99,7%), presisi sebesar 0,998 (99,8%), recall sebesar 0,997 (99,7) dan F1 sebesar 0,998 (99,8%).

#### Daftar Pustaka

- [1] Data Online Pusat Database BMKG. (Online), (<http://dataonline.bmkg.go.id/home>).
- [2] Pribadi, Khrisna S, dkk. 2008. Buku Pegangan Guru Pendidikan Siaga Bencana. Bandung : Pusat Mitigasi Bencana ITB.
- [3] R prathivi, *optimasi algoritme naive bayes untuk klasifikasi data gempa bumi di indonesia berdasarkan hiposentrum*, purwokerto : amikom, 2020.
- [4] BNPB.2014.Data & informasi Bencana Indonesia. (Online) ([http://dibi.bnppb.go.id/DesInventar/data\\_profil\\_wilayah.jsp](http://dibi.bnppb.go.id/DesInventar/data_profil_wilayah.jsp)) .
- [5] Mohammad Ihsan, *Analisa Ketahanan Gempa*. Bab. 2, FT UI, 2008.
- [6] Kholisul Fatikhin, *Data Mining-Pendahuluan*. Bab 1, FT UI, 2010.
- [7] Mochammad Haldi Widiyanto, *Algoritma Naives Baiyes*, Binus University, 2019.
- [8] Intan Noviantari Manyoe, Lantu Lantu, Samsu Arif, Rakhmat Jaya Lahay, *Earthquake Damage Level of Gorontalo Area Based on Seismicity and Peak Ground Acceleration*, Jambura Geoscience Review, 2019.
- [9] Dito Putro Utomo, Bister Purba. *Penerapan Datamining pada Data Gempa Bumi Terhadap Potensi Tsunami di Indonesia*. Prosiding Seminar Nasional Riset Information Science (SENARIS), 2019.
- [10] C. Fatichah, D. Purwitasari, *Deteksi Gempa Berdasarkan Data Twitter Menggunakan Decision ID3, Random Forest, dan SVM*. Jurnal Teknik ITS, 2017.
- [11] E. Buulolo, N. Silalahi, Fadlina and R. Rahim, *C4.5 Algorithm To Predict the Impact of the*, International Journal of Engineering Research & Technology (IJERT), vol. 6, no. 2, pp.10-15, 2017.
- [12] S. F. Rodiyansyah and E. Winarko, *Klasifikasi Posting Twitter Kemacetan Lalu Lintas Kota Bandung Menggunakan NaiveBayesian Classification*, Indonesian Journal of Computing and Cybernetics System (IJCCS), vol. 6, no. 1, pp. 91-100, 2012.



- [13] A. Jananto, *Algoritma Naive Bayes untuk Mencari Perkiraan Waktu Studi Mahasiswa*, Jurnal Teknologi Informasi DINAMIK , vol. 18, no. 1, pp. 09-16, 2013.
- [14] Sunarjo, M. T. Gunawan and S. Pribadi, *Gempa Bumi Edisi Populer*, Jakarta: Badan Meteorologi, Klimatologi dan Geofisika, 2012.
- [15] Rendra Dwi Lingga P., Chastine Fatchah, Diana Purwitasari, *Deteksi Gempa Berdasarkan Data Twitter Menggunakan Decision ID3, Random Forest, dan SVM*, Jurnal Teknik ITS, 2017.
- [16] S. F. Rodiyansyah and E. Winarko, *Klasifikasi Posting Twitter Kemacetan Lalu Lintas Kota Bandung Menggunakan NaiveBayesian Classification*, Indonesian Journal of Computing and Cybernetics System (IJCCS), vol. 6, no. 1, pp. 91-100, 2012.