

# **PENERAPAN *DATA MINING* UNTUK ANALISIS DATA BENCANA MILIK BNPB MENGGUNAKAN ALGORITMA *K-MEANS* DAN *LINEAR REGRESSION***

**Muhamad Iqbal Ramadhan<sup>1</sup>,  
Prihandoko<sup>2</sup>**

Fakultas Ilmu Komputer dan Teknologi Informasi, Universitas Gunadarma,  
<sup>1</sup>miiqbalrama@gmail.com  
<sup>2</sup>prihandoko@staff.gunadarma.ac.id

## ***Abstrak***

*Indonesia memiliki sejarah kejadian bencana alam yang cukup banyak, diantaranya adalah tsunami, gempa bumi, tanah longsor, kekeringan, banjir, letusan gunung berapi, dan sebagainya. Salah satu penyebab banyaknya potensi kejadian bencana alam di Indonesia adalah letak Indonesia yang berada di pertemuan lempeng – lempeng Eurasia, Indo-Australia dan Pasifik. Pertemuan lempeng dalam jangka panjang akan menghimpun energi yang suatu waktu akan lepas dan dapat menghasilkan bencana. Pengetahuan teknologi dan informasi pada saat ini sedang mengalami perkembangan yang pesat. Informasi tentang jumlah kejadian bencana alam dibutuhkan untuk penanggulangan bencana. Pengolahan data bencana alam yang umum dilakukan yaitu menggunakan teknik data mining, karena metode ini dianggap mampu menjadi solusi atas permasalahan penanggulangan bencana alam. Oleh karena itu, dalam penelitian ini membahas tentang pengelompokan jumlah data bencana dan prediksi data bencana yang akan terjadi 5 tahun kedepan menggunakan teknik data mining. Algoritma data mining yang digunakan adalah K-Means untuk clustering dan Linear Regression untuk prediksi data bencana.*

*Kata kunci: Clustering, Data Bencana Alam, Data Mining, Linear Regression, WEKA*

## **THE APPLICATION OF DATA MINING TO ANALYZE THE DISASTERS BY BNPB USING K-MEANS ALGORITHM AND LINEAR REGRESSION**

### ***Abstract***

*There are several disasters in the history of Indonesia, such as tsunami, earthquake, landslide, drought, flood, and mount eruption. One of those potential disasters are caused by the position of Indonesia which is in the central of slabs of Euro-Asian, Indo-Australian and Pasific. Those slabs, in a long period, will gather energy that could possibly result in disaster. Nowadays, the development of information and technology is capable of detecting the center of disaster. Information about numbers of disasters will be needed to avoid them. Generally, data processing of natural disasters is done by using data mining technique. This method is considered to be able give solution to the problem of natural disaster. Therefore, this study discusses a case of grouping the numbers of disasters and predict how many disasters could happen for next 5 years using data mining Technique. The Methods that will be employed are K-Means for clustering and Linear Regression for predicting the number of disasters.*

*Keywords : Clustering, Natural Disaster Data, Data Mining, Linear Regression, WEKA*

## PENDAHULUAN

Indonesia merupakan negara yang mempunyai struktur alam yang terdiri dari pertemuan lempeng-lempeng tektonik yaitu Eurasia, Indo-Australia dan Pasifik. Hal tersebut menjadikan kawasan Indonesia memiliki kondisi geologi yang sangat kompleks. Kondisi ini mengakibatkan banyak daerah-daerah di Indonesia yang sangat rawan terhadap bencana alam seperti gempa bumi, tsunami, serta letusan gunung berapi [1]. Daerah-daerah rawan bencana di Indonesia hampir semuanya berada pada daerah yang tingkat penduduknya sangat tinggi [2]. Selain bencana alam yang diakibatkan oleh faktor geologi, Indonesia juga sering mengalami bencana yang dipicu oleh kerusakan alam akibat perbuatan manusia seperti banjir dan tanah longsor. Berdasarkan data dari Badan Nasional Penanggulangan Bencana (BNPB), dalam kurun waktu 5 tahun saja yakni dari tahun 2004-2009, Indonesia harus mengalami 4.408 peristiwa bencana alam [1].

Saat ini pengetahuan teknologi dan informasi mengalami perkembangan yang sangat pesat. Teknologi yang semakin canggih membuat setiap orang mampu mengakses dan mendapat informasi secara cepat, tanpa mengenal batas-batas wilayah dan batasan waktu. Hal ini menyebabkan informasi menjadi sesuatu yang berharga dan sangat dibutuhkan untuk pengambilan keputusan. Salah satu informasi yang dibutuhkan oleh warga negara Indonesia adalah informasi mengenai kejadian bencana alam. Akan tetapi, informasi mengenai bencana alam yang tersaji saat ini masih bersifat acak, sulit dipahami dan belum terbukti keabsahannya. Walaupun bencana alam merupakan suatu kejadian yang tidak bisa dihindari, namun dampak bencana dapat dikurangi atau dapat diminimalisir dengan mengenali penyebab bencana dan mempelajari kejadian bencana yang telah

terjadi dengan menganalisa data bencana yang ada [3]. Selain itu, perlu dilakukan prediksi bencana yang akan terjadi di masa yang akan datang agar dapat menjadi indikator penanggulangan bencana alam yang akan terjadi. Pengolahan data menjadi sebuah informasi yang mudah dipahami sudah banyak dilakukan untuk berbagai kepentingan. Pengolahan data bencana alam yang umum dilakukan yaitu menggunakan teknik *data mining*.

Penggalian data (*data mining*) dapat didefinisikan sebagai proses menemukan pola dan tren yang tidak diketahui sebelumnya dalam basis data dan menggunakan informasi tersebut untuk membangun model prediktif [4]. *Data mining* berfungsi untuk proses pengambilan keputusan dari volume data yang besar yang disimpan dalam basis data, *data warehouse*, atau informasi yang disimpan dalam repositori [5]. Berdasarkan polanya *data mining* dikelompokkan menjadi deskripsi, estimasi, prediksi, klasifikasi, *clustering*, asosiasi [6]. *Data mining* merupakan inti dari proses *Knowledge Discovery in Database* (KDD) [7]. KDD adalah proses terorganisir untuk mengidentifikasi pola yang valid, baru, berguna, dan dapat dimengerti dari sebuah *dataset* yang besar dan kompleks [7].

Dewi Setianingsih dan RB Fajriya Hakim dalam jurnal yang berjudul "Penerapan *Data Mining* dalam Analisis Kejadian Tanah Longsor di Indonesia dengan Menggunakan *Association Rule* Algoritma *Apriori*" menggunakan *data mining* dengan metode *association rule* untuk mengolah data bencana tanah longsor di Indonesia karena ingin mengetahui informasi apa saja yang sering muncul bersamaan pada kejadian bencana tanah longsor [3]. Penelitian yang dilakukan oleh Tedy Rismawan dan Sri Kusumadewi menghasilkan pengelompokan mahasiswa berdasarkan nilai *Body Mass Index* (BMI) dan ukuran

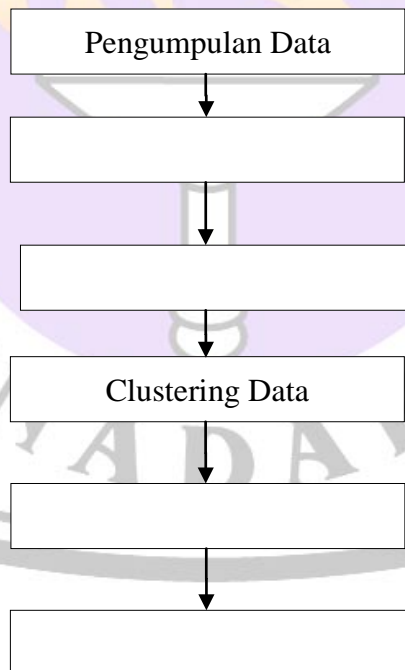
kerangka. Pengelompokan data dalam penelitian ini dilakukan dengan menggunakan metode *clustering K-Means* yaitu dengan mengelompokkan n buah objek ke dalam k kelas berdasarkan jaraknya dengan pusat kelas [8]. Penelitian yang dilakukan oleh Mirrah Zain, Ni Ketut Dewi Ari Jayanti, Yohanes Priyo Atmojo bertujuan untuk mengetahui jumlah pembukaan kelas di STIKOM Bali di periode berikutnya dengan cara meramalkan, dimana metode peramalan yang digunakan adalah metode Regresi Linear. Dengan mengimplementasikan regresi linear ke dalam sistem pembukaan kelas ini memberikan kemudahan dalam melakukan perhitungan serta mengurangi terjadinya drop kelas [9].

Berdasarkan latar belakang masalah tersebut, pada penelitian ini diterapkan salah satu teknik *data mining* menggunakan algoritma *K-Means* dan

algoritma *Linear Regression* untuk analisis data bencana alam milik Badan Nasional Penanggulangan Bencana (BNPB) pada WEKA. Algoritma *K-Means* digunakan untuk melakukan pengelompokan (*clustering*) data bencana alam dan algoritma *Linear Regression* untuk melakukan prediksi data bencana alam yang akan terjadi di masa yang akan datang.

## METODE PENELITIAN

Penelitian ini bertujuan untuk mengelompokkan jumlah data bencana alam milik BPNB dan prediksi data bencana yang akan terjadi 5 tahun kedepan menggunakan teknik *data mining*. Adapun keseluruhan tahapan dalam penelitian ini ditunjukkan oleh bagan pada Gambar 1.



Gambar 1. Diagram Alur Metodologi Penelitian



Pada penelitian ini dimulai dengan melakukan pengumpulan data. Data pada penelitian ini berupa data bencana alam 34 provinsi di Indonesia dari tahun 2005 sampai tahun 2015, yang diperoleh dari BNPB. Data-data yang telah didapatkan pada tahap pengumpulan data kemudian dilakukan transformasi pada data-data yang berjenis data nominal, yaitu seperti nama provinsi dan jenis bencana alam. Sebelum dilakukan tahap *clustering* dilakukan *pre-processing* data, dimana data-data yang berjenis data nominal tersebut diinisialisasikan ke dalam bentuk angka.

Setelah semua data ditransformasi ke dalam bentuk angka, maka data-data tersebut telah dapat dikelompokkan dengan menggunakan algoritma *K-Means clustering*. Data bencana ini dikelompokkan berdasar kemiripan karakteristik dari setiap data, sehingga dapat ditemukan informasi yang tersembunyi dari data-data tersebut. Semua data yang telah dikelompokkan tadi kemudian diolah lagi untuk mendapatkan hasil prediksi data bencana yang akan terjadi 5 tahun kedepan menggunakan teknik *data mining*. Prediksi data bencana ini menggunakan algoritma *Linier Regression*. Berdasarkan analisa hasil dan pengujian data dengan menerapkan algoritma *K-Means clustering* dan algoritma *Linier Regression* ini diperoleh kesimpulan dari penelitian yang telah dilakukan.

## HASIL DAN PEMBAHASAN

### **Pre-Process Data Bencana 34 Provinsi di Indonesia**

Pada tahap dilakukan *Pre-Process (Cleaning)* data bencana 34 provinsi di Indonesia. *Pre-Process* dapat dilakukan dengan *import* terlebih dahulu *dataset* bencana alam ke WEKA, kemudian setelah berhasil *import* dapat melakukan proses *cleaning* dengan

mengetikkan perintah berikut pada *Filter chooserSectionWEKA* :

```
weka.filters.unsupervised.attribute
.Remove -R 2, 3, 4, 6, 7, 8, 9, 10,
11, 12, 13, 14, 16, 17, 18, 19, 20
```

Kemudian langkah selanjutnya yaitu dengan mengkonversi atribut yang ada dari *numeric* menjadi nominal dengan mengetikkan perintah berikut pada *Filter chooserSectionWEKA* :

```
weka.filters.unsupervised.attribute.
NumericToNominal -R 1
```

### **Clustering Data Bencana 34 Provinsi di Indonesia**

Pada tahap ini dilakukan *clustering* data bencana menggunakan algoritma *K-Means*. Langkah yang dilakukan adalah dengan mengetikkan perintah berikut pada *Clusterer chooser Section WEKA* :

```
weka.Clusterers.SimpleKMeans -init 2
-max-candidates 100 -periodic -
pruning 10000 -min-density 2.0 -t1 -
1.25 -t2 -1.0 -N 11 -A
"weka.core.EuclideanDistance-R first-
last" -I 500 -num-slots 1 -S 10
```

### **Prediksi Data Bencana 34 Provinsi di Indonesia**

Pada tahap ini dilakukan prediksi data bencana menggunakan fitur *Forecast* dari WEKA yang memanfaatkan algoritma *Linear Regression*. Prediksi bencana selama 5 tahun kedepan yaitu mulai dari tahun 2016 sampai dengan tahun 2020. Hal tersebut dapat dilakukan dengan mengetikkan perintah berikut pada *Base Learner Configuration Chooser Section WEKA* :

```
weka.Classifiers.functions.LinearRegression
-S 0 -R 1.0E-8 -num-decimal-places 4.-
F (Jenis Bencana yang akan di prediksi) -
L 1-M (Banyaknya tahun yang diprediksi)-
G Tahun
```

Setelah menetikkan perintah tersebut, maka WEKA menampilkan jumlah bencana yang akan terjadi pada provinsi dan jenis bencana yang sedang dikelompokkan selama tahun 2016 sampai dengan tahun 2020.

### Analisa Hasil Pengujian Data Bencana 34 Provinsi di Indonesia

Pada tahap ini dilakukan analisa dari hasil pengujian, yaitu *clustering* dan prediksi data bencana 34 Provinsi di Indonesia. Hasil *clustering* dan prediksi data bencana per Provinsi di Indonesia dapat dilihat pada Tabel 1.

Berdasarkan data yang ada dalam Tabel 1, diperoleh hasil mining data bencana di seluruh provinsi di Indonesia sebagai berikut :

1. Provinsi Jawa Tengah memiliki jumlah bencana terbesar, yaitu sebanyak 3055 kejadian dalam kurun waktu 2005-2015 dengan bencana yang paling sering terjadi adalah bencana tanah longsor sebanyak 174 kejadian.
2. Berdasarkan prediksi menggunakan algoritma *Linear Regression*, Provinsi Jawa Barat memiliki jumlah bencana terbesar, yaitu sebanyak 3639 kejadian dengan bencana yang paling sering terjadi adalah tanah

longsor sebanyak 273 kejadian.

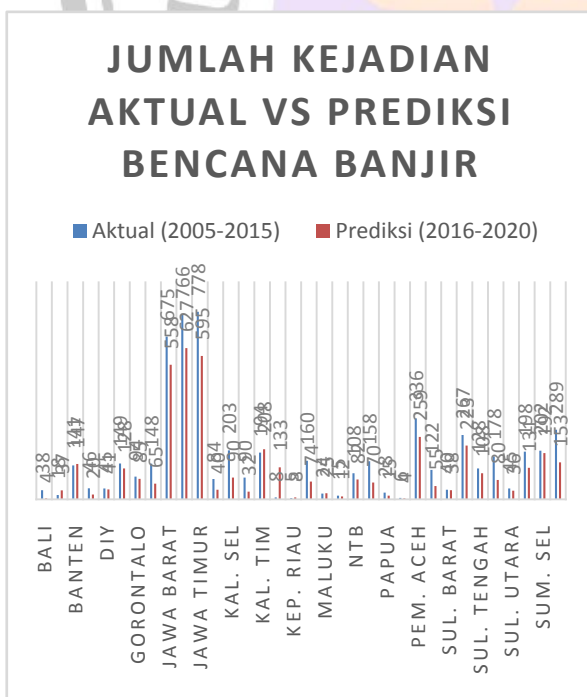
3. Berdasarkan hasil prediksi untuk periode tahun 2016 – 2020, Banjir akan tetap menjadi bencana yang mendominasi 20 provinsi di Indonesia, yaitu sebesar 4375 kejadian dengan probabilitas sebesar 0.76 kali berdasarkan data bencana periode sebelumnya yaitu 5786 kejadian bencana banjir atau sama dengan 26 provinsi di Indonesia.
4. Berdasarkan hasil prediksi untuk periode tahun 2016 – 2020, akan terjadi 15023 bencana di seluruh Provinsi di Indonesia dengan probabilitas sebesar 0.99 kali berdasarkan data bencana periode sebelumnya yaitu 15133 kejadian bencana

Pada penelitian ini juga dilakukan perbandingan data bencana per Provinsi di Indonesia. Perbandingan dilakukan dengan membandingkan data bencana alam yang sejenis pada kejadian actual yaitu tahun 2005 – 2015 yang dihasilkan dari algoritma *K-Means* dengan data prediksi yang dihasilkan dari algoritma *Linear Regression*. Adapun hasil dari perbandingan ini ditunjukkan oleh Gambar 2

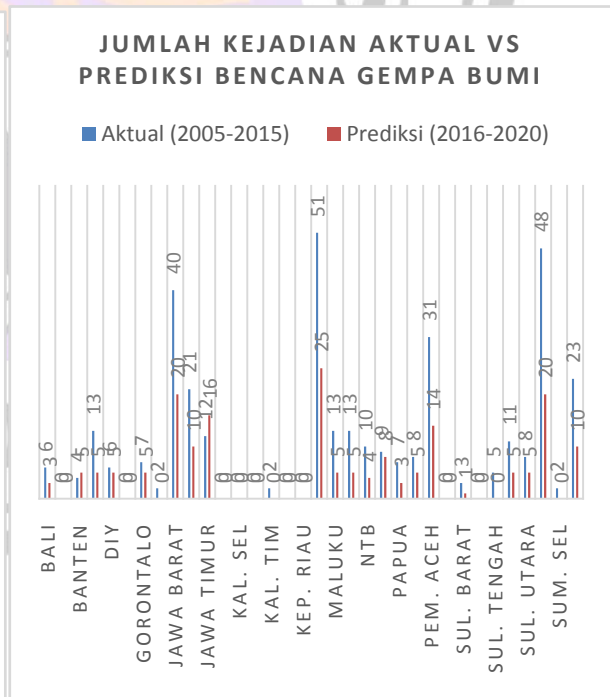
Tabel 1. Hasil Clustering dan Prediksi Data Bencana per Provinsi di Indonesia

No.	Provinsi	Jumlah Kejadian Aktual (2005-2015)	Jumlah Kejadian Prediksi (2016-2020)	Probabilitas Peningkatan Terjadinya Bencana	Kejadian Paling Sering Terjadi (Aktual)	Kejadian Paling Sering Terjadi (Prediksi)
1	Bali	197	80	0.41	Puting Beliung	Puting Beliung
2	Bangka Belitung	56	109	1.95	Puting Beliung	Puting Beliung
3	Banten	329	385	1.17	Banjir	Banjir
4	Bengkulu	97	45	0.46	Banjir	Banjir
5	DI Yogyakarta	254	443	1.74	Puting Beliung	Puting Beliung
6	DKI Jakarta	174	151	0.87	Banjir	Banjir
7	Gorontalo	138	129	0.93	Banjir	Banjir
8	Jambi	252	145	0.58	Banjir	Banjir
9	Jawa Barat	2462	3639	1.48	Tanah Longsor	Tanah Longsor
10	Jawa Tengah	3055	2816	0.92	Tanah Longsor	Tanah Longsor

11	Jawa Timur	1941	1837	0.95	Banjir	Tanah Longsor
12	Kalimantan Barat	132	69	0.52	Banjir	Banjir
13	Kalimantan Selatan	379	286	0.75	Banjir	Puting Beliung
14	Kalimantan Tengah	111	48	0.43	Banjir	Banjir
15	Kalimantan Timur	328	638	1.95	Banjir	Puting Beliung
16	Kalimantan Utara	8	133	16.63	Banjir	Banjir
17	Kepulauan Riau	21	26	1.24	Puting Beliung	Puting Beliung
18	Lampung	317	145	0.46	Banjir	Banjir
19	Maluku	101	147	1.46	Banjir	Banjir
20	Maluku Utara	52	32	0.62	Banjir	Banjir
21	Nusa Tenggara Barat	269	250	0.93	Banjir	Banjir
22	Nusa Tenggara Timur	500	301	0.60	Puting Beliung	Puting Beliung
23	Papua	76	120	1.58	Banjir	Banjir dan Tanah Longsor
24	Papua Barat	19	12	0.63	Gempa Bumi	Gempa Bumi
25	Pemerintah Aceh	652	391	0.60	Banjir	Banjir
26	Riau	242	301	1.24	Banjir	Puting Beliung
27	Sulawesi Barat	90	79	0.88	Banjir	Banjir
28	Sulawesi Selatan	603	447	0.74	Banjir	Banjir
29	Sulawesi Tengah	192	150	0.78	Banjir	Banjir
30	Sulawesi Tenggara	437	218	0.50	Banjir	Banjir
31	Sulawesi Utara	161	137	0.85	Banjir	Banjir
32	Sumatera Barat	568	566	1.00	Banjir	Tanah Longsor
33	Sumatera Selatan	410	364	0.89	Banjir	Banjir
34	Sumatera Utara	510	384	0.75	Banjir	Banjir
	Grand Total	15133	15023	0.99		

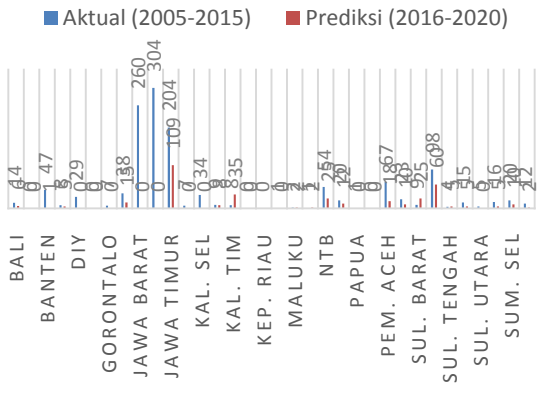


(a)



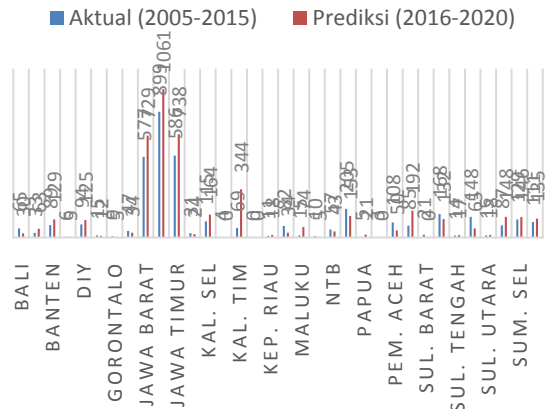
(b)

### JUMLAH KEJADIAN AKTUAL VS PREDIKSI BENCANA KEKERINGAN



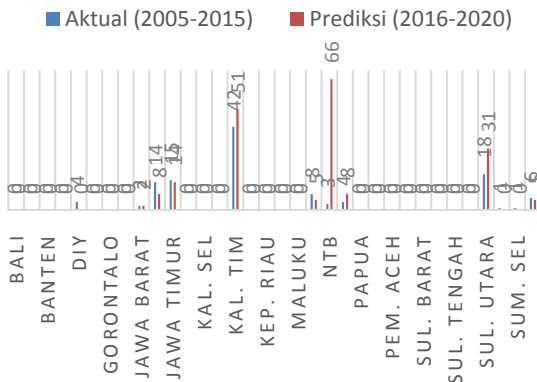
(c)

### JUMLAH KEJADIAN AKTUAL VS PREDIKSI BENCANA PUTING BELIUNG



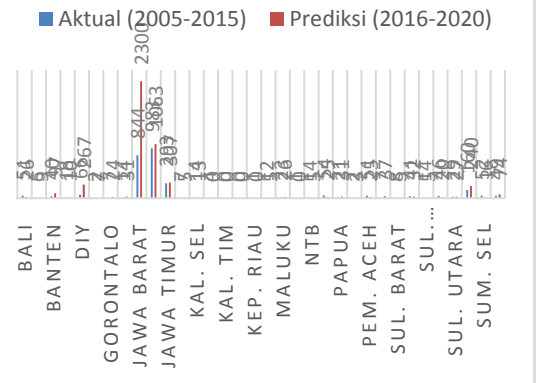
(e)

### JUMLAH KEJADIAN AKTUAL VS PREDIKSI BENCANA LETUSAN GUNUNG API



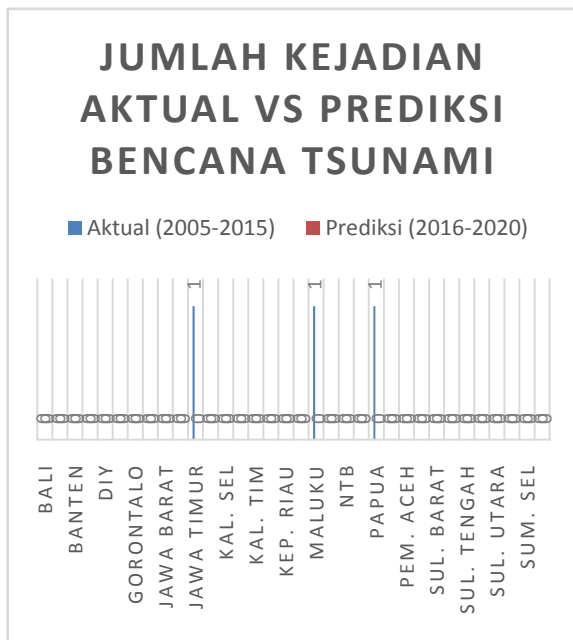
(d)

### JUMLAH KEJADIAN AKTUAL VS PREDIKSI BENCANA TANAH LONGSOR



(f)





(g)

Gambar 2. Perbandingan Jumlah Bencana Sejenis pada Kejadian Aktual dengan Prediksi Bencana (a) Bencana Banjir (b) Bencana Gempa Bumi (c) Bencana Kekeringan (d) Bencana Letusan Gunung Api (e) Bencana Puting Beliung (f) Bencana Tanah Longsor (g) Bencana Tsunami

## KESIMPULAN DAN SARAN

Pada penelitian ini telah berhasil dilakukan pengelompokan jumlah data bencana seluruh provinsi di Indonesia milik BNPD dan prediksi data bencana yang akan terjadi 5 tahun kedepan menggunakan teknik *data mining* dengan menerapkan Algoritma *K-Means* dan algoritma *Linear Regression*. Algoritma *K-Means* digunakan untuk mengelompokkan data bencana per provinsi dalam kurun waktu 2005-2015. Berdasarkan algoritma ini diperoleh kesimpulan bahwa Provinsi Jawa Tengah memiliki jumlah bencana tanah longsor terbesar. Selanjutnya berdasarkan algoritma *Linear Regression* untuk prediksi data bencana periode tahun 2016 – 2020 diperoleh hasil bahwa Provinsi Jawa Barat memiliki jumlah bencana tanah longsor

terbesar. Berdasarkan hasil prediksi juga diperoleh hasil bahwa bencana banjir akan tetap menjadi bencana yang mendominasi 20 provinsi di Indonesia.

Penelitian ini masih memiliki kekurangan dikarenakan dalam penelitian ini hanya berfokus pada jumlah kejadian bencana di 34 Provinsi di Indonesia, sehingga hasil prediksi belum dapat dibuktikan keabsahannya. Untuk penelitian selanjutnya pada dataset sebaiknya dilengkapi dengan beberapa indikator seperti curah hujan ataupun tingkat kelembaban yang telah dikelompokkan dalam kelas tertentu, sehingga hasil prediksi data bencana lebih akurat.

## DAFTAR PUSTAKA

- [1] Indriasari, Th.D. dan Sidhi, T.A.P. 2011. *Sistem Pencarian Orang Hilang Berbasis Mobile Web dengan Social Network Analysis*. Seminar Nasional Informatika 2011 (semnasIF 2011), Veteran Yogyakarta. pp. E12–E19.
- [2] Bachtiar, A., waktu akses Juni 2011, Bencana Nasional Gempa-Tsunami Aceh: "Tuntutan untuk lebih Serius dalam Mitigasi Bencana", <http://www.itb.ac.id/news/381.xhtml>.
- [3] Setianingsih, D. dan Hakim, R.B.F. 2015. *Penerapan Data Mining dalam Analisis Kejadian Tanah Longsor di Indonesia dengan Menggunakan Association Rule Algoritma Apriori*. Prosiding Seminar Nasional Matematika dan Pendidikan Matematika UMS. pp. 731–741.
- [4] Iswari, N.M.S. 2011. "Penggunaan Teknik *Data Mining* untuk Manajemen Resiko Sistem Informasi Rumah Sakit". *ULTIMATICS*. Vol. 3, No. 2, pp. 16–22.
- [5] Han, J., dan Kamber, M. 2006. *Data Mining Concept and Tehniques*. San Fransisco: MorganKauffman.



- [6] Larose, D. T. 2005. *Discovering Knowledge in Data*. New Jersey: John Willey & Sons, Inc.
- [7] Maimon, O., dan Rokach, L. 2005. *Data Mining and Knowledge Discovey Handbook*. NewYork: Springer.
- [8] Rismawan, T. dan Kusumadewi, S. 2008. *Aplikasi K-Means untuk Pengelompokkan Mahasiswa Berdasarkan Nilai Body Mass Index (BMI) & Ukuran Kerangka*. Seminar Nasional Aplikasi Teknologi Informasi 2008 (SNATI 2008), Yogyakarta. pp. E43–E48.
- [9] Zain, M., Jayanti, D.A., dan Atmojo, Y.P. 2014. "Implementasi *Forecasting* pada Perancangan Sistem Pembukaan Kelas di STIKOM Bali dengan Megggunakan Metode Regresi Linier". *EKSPLORA INFORMATIKA*. Vol. 3, No. 1, pp. 17–28.

