

**PENERAPAN ALGORITMA KLASIFIKASI NAIVE BAYES
UNTUK DATA STATUS HUNI RUMAH BANTUAN DANA
REHABILITASI DAN REKONSTRUKSI PASCA BENCANA ERUPSI
GUNUNG MERAPI 2010**

**IMPLEMENTATION OF NAIVE BAYES CLASSIFICATION
ALGORITHM FOR OCCUPANCY HOUSE DATA STATUS FUND
ASSISTANCE REHABILITATION AND RECONSTRUCTION POST
OF ERUPTION DISASTERS MERAPI MOUNTAIN 2010**

Nurhadi Wijaya

Program Studi Teknik Informatika, Fakultas Sains dan Teknologi Universitas Respati Yogyakarta
adhv.respati@yahoo.com

Abstrak

Bencana Erupsi gunung Merapi berikut susulan material lahar hujan pada Tahun 2010 mengakibatkan kerusakan rumah dan infrastruktur di wilayah Kabupaten Sleman D.I.Yogyakarta dan Kabupaten Magelang Jawa Tengah. Melalui Perka BNPB No.5 Tahun 2011, pemerintah menginstruksikan rencana dan aksi rehabilitasi dan rekonstruksi pasca erupsi dilakukan dengan skema program Rehabilitasi dan Rekonstruksi Masyarakat dan Permukiman Berbasis Masyarakat. Skema program ini telah membangun rumah sebanyak 2.516-unit bagi warga yang terdampak erupsi Merapi dan lahar hujan. Menurut *Key Performance Indikator (KPI) The World Bank*, status huni rumah terbangun merupakan salah satu indikator kinerja program rehab rekon. Pelaksanaan program rehab dan rekon ini sebagian besar didokumentasikan secara digital dan terekam ke dalam basis data. Dalam Ilmu Teknologi Informasi dibidang data mining, basis data merupakan aset yang dapat digunakan sebagai bahan pengenalan dan penemuan pola-pola data yang dapat dipelajari dan diteliti guna menyelesaikan permasalahan. Basis data yang dimiliki Satker rehab rekon merekam sebanyak 2.146-unit rumah huni sudah dihuni dan 370 rumah belum dihuni. Hasil penelitian/eksperimen menunjukkan bahwa penerapan algoritma klasifikasi Naive Bayes dapat diterapkan terhadap data status huni rumah bantuan dana rehabilitasi dan rekonstruksi pasca erupsi Merapi 2010 dengan hasil nilai akurasi klasifikasi mencapai sebesar 89,59% dan nilai performa klasifikasi AUC mencapai 0,826

Kata kunci : Erupsi Merapi, Data Mining, Naive Bayes, Klasifikasi, Rehab Rekon, Status huni

Abstract

Disaster Eruption of Mount Merapi and the following a mixture of lava rain material in 2010 resulted in damage to homes and infrastructure in the Sleman Regency of D.I.Yogyakarta and Magelang District of Central Java. Through Perka BNPB No.5 of 2011, the government instructed plans and actions for rehabilitation and reconstruction after the eruption was carried out with the scheme of the Community Rehabilitation and Reconstruction and Community Based Settlement program. The program scheme has built 2,516-unit houses for residents affected by Merapi and rain lava eruptions. According to The World Bank's Key Performance Indicator (KPI), the occupancy status of built houses is one of the indicators of the performance of the rehabilitation and reconstruction program. The implementation of the rehabilitation and reconstruction program is mostly digitally documented and recorded in the database. In Information Technology in the field of data mining, the database is an asset that can be used as an introduction and discovery of data patterns that can be studied and researched to solve problems. The database owned by the reconstruction rehabilitation work unit recorded 2,146 housing units has been occupied and 370 houses have not been occupied. The results of the research / experiment show that the application of the Naive Bayes classification algorithm can be applied to the occupancy status data of houses

for rehabilitation and reconstruction assistance after the 2010 Merapi eruption with the classification accuracy reaching 89.59% and the AUC classification performance value reaching 0.826

Keywords: Merapi Eruption, Data Mining, Naive Bayes, Classification, Reconstruction Rehabilitation, Occupied status

1. PENDAHULUAN

Bencana erupsi gunung Merapi dan Lahar Hujan pada tahun 2010 telah menyebabkan kerusakan di Wilayah Kabupaten Sleman Yogyakarta dan di Wilayah Kabupaten Magelang Provinsi Jawa Tengah. Melalui peraturan kepala (PerKa) Badan Nasional Penanggulangan Bencana (BNPB) Nomor 5 Tahun 2011 tentang rencana aksi rehabilitasi dan rekonstruksi (Renaksi) daerah setelah bencana Erupsi Merapi di 2010, menggunakan skema rehabilitasi dan rekonstruksi perumahan dan permukiman dengan komunitas berbasis masyarakat. Skema rehabilitasi dan rekonstruksi dimaksudkan tidak hanya untuk memindahkan pemukiman secara fisik tetapi juga untuk memindahkan kehidupan dan mata pencaharian mereka. Membuat pemukiman di tanah baru dilakukan melalui pendekatan pemberdayaan masyarakat dengan mempromosikan kombinasi pembangunan dan nilai berbasis masyarakat [1].

Berdasarkan indikator kinerja skema program rehabilitasi dan rekonstruksi permukiman dengan komunitas berbasis masyarakat yang dikeluarkan oleh Bank Dunia, status huni rumah adalah salah satu indikator kinerja dari keberhasilan kinerja skema program rehabilitasi dan rekonstruksi tersebut. Semakin banyak rumah yang dihuni, semakin baik kinerjanya [2].

Penelitian terkait bencana pasca erupsi Merapi, khususnya data penerima dana bantuan rekonstruksi dan rehabilitasi perumahan pasca bencana Merapi 2010 menggunakan pendekatan Ilmu Teknologi Informasi pada bidang *data mining* dengan teknik klasifikasi belum banyak dilakukan. Terutama penelitian yang berkaitan langsung dengan status hunian rumah terbangun melalui program rehabilitasi dan rekonstruksi tidak mudah ditemukan.

Data status hunian yang dimiliki oleh satuan unit kerja (satker) rehabilitasi dan rekonstruksi pasca bencana Erupsi gunung Merapi 2010, merekam sebanyak 2.146 unit rumah telah dihuni, sebanyak 370 rumah belum dihuni atau belum diketahui statusnya.

Berdasarkan status kelas yang sudah dihuni dan belum dihuni tersebut, mendasari dan membangkitkan sebuah pertanyaan berikut keinginan untuk meneliti bagaimana menerapkan pendekatan Teknologi Informasi di bidang data mining dengan teknik klasifikasi algoritma Naive Bayes untuk mengklasifikasikan data status huni rumah bantuan dana rehab rekon pasca bencana erupsi gunung Merapi Tahun 2010.

2. DASAR TEORI

2.1 Data Mining

Data mining merupakan suatu proses yang memiliki tujuan untuk menemukan pola dari data yang sudah ada di dalam basis data agar dapat dimanfaatkan untuk menyelesaikan masalah, mengelompokkan objek atau data berdasarkan kemiripan data, sehingga anggota dalam kelompok memiliki banyak kemiripan dibandingkan dengan kelompok lain. Teknik klasifikasi dalam data mining merupakan teknik pembelajaran yang digunakan untuk dapat memprediksi nilai dari atribut kategori target. Klasifikasi bertujuan untuk membagi objek yang ditugaskan hanya ke salah satu nomor kategori yang disebut kelas (*Class*) [3].

2.2 Klasifikasi

Menurut [4], klasifikasi adalah proses penemuan model (atau fungsi) yang menggambarkan dan membedakan kelas data atau konsep yang bertujuan agar bisa digunakan untuk memprediksi kelas dari objek yang label kelasnya tidak diketahui.

Terdapat dua langkah tahapan proses klasifikasi. Tahap pertama adalah learning (*fase training*), dimana algoritma klasifikasi dibuat agar dapat menganalisa data *training* lalu direpresentasikan ke dalam bentuk *rule* klasifikasi. Tahap kedua adalah klasifikasi, dimana data tes digunakan untuk memperkirakan akurasi dari *rule* klasifikasi Metode yang dapat digunakan untuk mengukur akurasi algoritma klasifikasi, antara lain adalah: *cross validation*, *confusion matrix*, dan kurva ROC (*Receiver Operating Characteristic*).

2.3 Algoritma Naive Bayes

Naive Bayes Classifier disingkat (NBC) yang sering disebut juga *Bayesian Classification* adalah metode pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan dari suatu *Class*. NBC berlandaskan pada teorema Bayes yang memiliki kemampuan klasifikasi seperti *decision tree* dan *neural network*. NBC terbukti memiliki akurasi dan kecepatan tinggi saat diaplikasikan ke dalam database yang besar [5]. Algoritma Naive Bayes merupakan salah satu algoritma yang terdapat pada teknik klasifikasi. Naive Bayes merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan asal Inggris Thomas Bayes, yaitu memprediksi peluang dimasa depan berdasarkan pengalaman dimasa sebelumnya sehingga dikenal sebagai Teorema Bayes. Teorema tersebut dikombinasikan dengan Naive Bayes dimana diasumsikan kondisi antar atribut saling bebas. Klasifikasi Naive Bayes diasumsikan bahwa ada atau tidak ciri tertentu dari sebuah kelas tidak ada hubungannya dengan ciri dari kelas lainnya [6]. Adapun Persamaan teorema Bayes adalah sebagai berikut :

$$P(H | X) = \frac{P(X | H) \cdot P(H)}{P(X)} \quad (1)$$

Keterangan:

X : Data dengan *class* yang belum diketahui P(X) : Probabilitas X

H : Hipotesis data X merupakan suatu *class* spesifik

P(H|X) : Probabilitas hipotesis berdasar kondisi (posteriori probability)

P(H) : Probabilitas hipotesis H (prior probability)

P(X|H) : Probabilitas X berdasarkan kondisi pada hipotesis H

Sedangkan untuk pengklasifikasian dengan data kontiyu digunakan persamaan Densitas Gauss :

$$P(X_i = x_i | Y = Y_j) = \frac{1}{\sqrt{2\pi\sigma_{ij}}} e^{-\frac{(x_i - \mu_{ij})^2}{2\sigma_{ij}^2}} \quad (2)$$

Keterangan:

P : Peluang μ : Mean, menyatakan rata – rata dari seluruh atribut

x_i : Nilai Atribut ke i Y : Kelas yang dicari X_i : Atribut ke i

σ : Deviasi standar, menyatakan varian dari seluruh atribut Y_j : Sub Kelas Y yang dicari

2.4 Cross Validation

Cross Validation adalah teknik validasi dengan cara membagi data secara acak ke dalam k bagian dan masing-masing bagian akan dilakukan proses klasifikasi. *Cross*

validation akan melakukan percobaan sebanyak k . Secara umum pengujian nilai k dilakukan sebanyak 10 kali untuk memperkirakan akurasi estimasi [7]. Tiap percobaan akan menggunakan satu data testing dan $k-1$ bagian akan menjadi data *training*, kemudian data testing itu akan ditukar dengan satu buah data *training* sehingga untuk tiap percobaan akan didapatkan data testing yang berbeda-beda. Data *training* adalah data yang akan dipakai dalam melakukan pembelajaran sedangkan data *testing* adalah data yang belum pernah dipakai sebagai pembelajaran dan akan berfungsi sebagai data penguji kebenaran atau keakuratan hasil pembelajaran.

2.4 Confusion Matrix

Confusion matrix adalah salah satu alat ukur berbentuk matrik 2×2 yang digunakan untuk mendapatkan jumlah ketepatan klasifikasi dataset terhadap kelas aktif dan tidak aktif pada algoritma yang dipakai. Evaluasi model klasifikasi dilandaskan pada pengujian untuk memperkirakan obyek yang benar dan salah. Urutan pengujian ditabulasikan dalam *confusion matrix* dimana *class* yang diprediksi ditampilkan dibagian atas matriks dan *class* yang diamati disisi kiri. Setiap sel memiliki isi angka yang menunjukkan berapa banyak kasus yang sebenarnya dari kelas yang diamati untuk diprediksi [8]. *Confusion matrix* memberikan keputusan yang diperoleh dalam training dan testing, *confusion matrix* memberikan penilaian dalam performance klasifikasi berdasarkan objek dengan benar atau salah [3]. Berikut di bawah ini merupakan tabel *confusion matrix*:

Tabel 1 Confusion Matrix

Classification	Predicted Class		
	Class = Yes	Class = No	
Observed Class	Class = Yes	a (True Positive-TP)	b (False Negative-FN)
	Class = No	c (False Positive-FP)	d (True Negative-TN)

Keterangan:

TP = Prediksi positif yang positif

FN = Prediksi positif yang negatif

FP = Prediksi negatif yang positif

TN = Prediksi negatif yang negatif

Untuk perhitungan nilai dari hasil klasifikasi pada penelitian ini akan dihitung nilai *accuracy* hasil klasifikasi dengan persamaan perhitungan sebagai berikut:

$$Accuracy = \left(\frac{a+d}{a+b+c+d} \right) = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

2.5 ROC (Receiver Operating Characteristic) Curve

Kurva ROC dibagi dalam dua dimensi, dimana tingkat TP diplot pada sumbu Y dan tingkat FP diplot pada sumbu X. Tetapi untuk merepresentasikan grafis yang menentukan klasifikasi mana yang lebih baik, digunakan metode yang mengbitung luas daerah di bawah kurva ROC yang disebut AUC (*Area Under the ROC curve*) yang diartikan sebagai probabilitas [7]. AUC mengukur kinerja diskriminatif dengan memperkirakan probabilitas *output* dari sampel yang di pilih secara acak dari populasi positif atau negatif, semakin besar AUC, semakin kuat klasifikasi yang digunakan. Karena AUC adalah bagian dari daerah unit persegi nilainya akan selalu antara 0,0 dan 1,0.

3. METODOLOGI PENELITIAN

Penelitian ini menggunakan penelitian studi literatur/pustaka dan penelitian percobaan (*Experimental*). Penelitian studi literatur atau studi pustaka dilakukan untuk menambah khasanah ilmu pengetahuan sebelum melakukan *experiment* agar tidak melebar dan fokus terhadap penelitian yang dilakukan. Penelitian ini melibatkan analisa data serta perlakuan pada dataset dengan memperhatikan parameter, atribut, variabel termasuk cara uji/testing yang digunakan.

3.1 Dataset

Dataset yang digunakan pada penelitian ini menggunakan data *Secondary Private Database Management Information System (MIS)* rehab rekon pasca erupsi Merapi dan lahar hujan Tahun 2010 yang dikumpulkan dan diperoleh melalui instansi Satuan Kerja (Satker) rehab rekon pasca Erupsi Merapi 2010. Jumlah *record* atau *instances* adalah sebanyak 2.516, dan Atribut yang digunakan sebanyak 11 Atribut. Data yang digunakan ini berupa data set status hunian (sudah dihuni dan belum dihuni) rumah bantuan dana rehab rekon. Adapun analisa dataset yang digunakan dapat dilihat pada tabel berikut ini :

Tabel 2 Aanalisa dataset

No.	Atribut	Keterangan
1.	Anggaran	Data <i>Polynomial</i> , memiliki Entitas: BNPB 2011; JRF 2011; PSF 2011; PSF 2012; PSF 2013
2.	JenisKelamin	Data <i>Binominal</i> , memiliki entitas: Laki-Laki dan Perempuan
3.	JumlahAnggotaRumahTangga	Data <i>Polynomial</i> . Dengan entitas : Kecil (1-4 orang); Sedang (5-7 orang); Besar (lebih dari 7 orang)
4.	PenghuniPerempuan	Data Kontinyu – <i>Numeric-Integer</i>
5.	PekAir	Data <i>Binominal</i> . Memiliki entitas Sudah dan Belum
6.	PekListrik	Data <i>Binominal</i> . Memiliki entitas: Sudah dan Belum
7.	TingkatKerusakan	Data <i>Binominal</i> memiliki entitas : Hilang/Rusak Berat & Sedang/Rusak Ringan
8.	SwadayaMasyarakat	Data <i>Binominal</i> dengan entitas : $\geq 10\%$ & $< 10\%$
9.	StatusLahan	Data <i>Polynomial</i> . memiliki entitas : Mandiri; Mandiri kolektif; TKD
10.	LuasBangunan	Data <i>Polynomial</i> yang memiliki entitas: 36 m²; > 36 m²; < 36 m²
11.	StatusHuni	Data <i>Binominal</i> yang memiliki entitas. Sudah & Belum , data ini Merupakan <i>Class</i> dari dataset yang digunakan.

3.2 Populasi, Sampel dan metode/Teknik penentuan sampel

Penelitian ini menggunakan sample dengan tipe desain sampling *probability Sampling*, dimana setiap elemen populasi mempunyai kesempatan yang sama untuk diseleksi sebagai subyek dalam sampel. Sedangkan teknik penentuan sampel menggunakan *Stratified Random Sampling*. Teknik ini digunakan untuk mengurangi pengaruh faktor heterogen dan melakukan pembagian elemen-elemen populasi ke dalam strata. Kemudian pada masing – masing strata dipilih sampel secara random sesuai proporsinya.

3.3 Material dan alat yang digunakan

Dalam melakukan eksperimen uji coba/test terhadap data status huni rumah bantuan dana rehab rekon pasca erupsi Merapi 2010 ini digunakan spesifikasi perangkat keras dan perangkat lunak yang dapat dilihat pada Tabel 3 berikut ini:

Tabel 3 Spesifikasi perangkat lunak dan keras penelitian

Perangkat Lunak	Perangkat Keras
Sistem Operasi <i>Windows 7</i>	<i>Processor</i> : Intel Core i5
Microsoft <i>Office Spreadsheet - Excel 2016</i>	Memory (RAM) : 4GB
<i>Rapid Miner 5.0.3.0.15</i>	Hard Drive Disk : 500 GB

3.4 Evaluasi dan Uji validitas

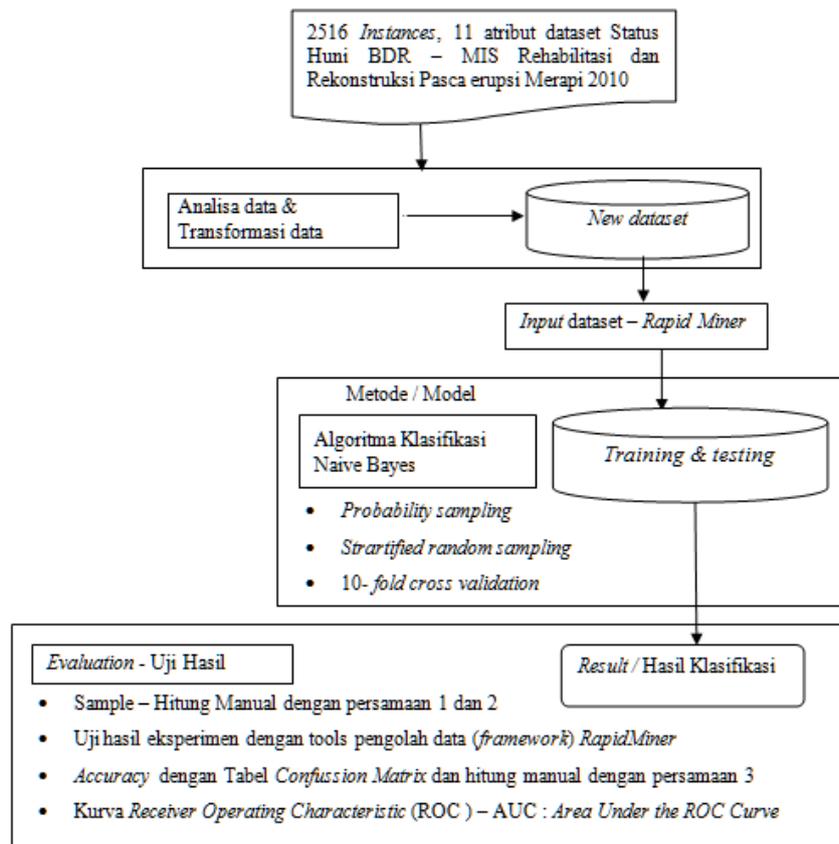
Selain dilakukan penghitungan secara manual terhadap sampel data dengan menggunakan persamaan (1) dan (2), evaluasi dan validasi pengujian dataset status huni bantuan dana rumah rehab rekon dengan algoritma klasifikasi *Naive Bayes Classification* (NBC) pada eksperimen ini, digunakan validasi *10-fold cross validation* dengan *Stratified Random Sampling*. Untuk menguji akurasi digunakan tabel *confusion matrix* dan Kurva ROC – AUC. Kurva AUC mengukur kinerja dengan memperkirakan probabilitas output dari sampel yang dipilih secara acak dari populasi positif atau negatif, semakin besar AUC, semakin kuat klasifikasi yang digunakan. Performa keakurasian AUC dapat diklasifikasikan menjadi lima kelompok [7], yaitu:

- 0.90 - 1.00 = klasifikasi sangat baik (*excellent classification*).
- 0.80 - 0.90 = klasifikasi baik (*good classification*).
- 0.70 - 0.80 = klasifikasi cukup (*fair classification*).
- 0.60 - 0.70 = klasifikasi buruk (*poor classification*).
- 0.50 - 0.60 = klasifikasi salah (*failure*).

Eksperimen atau uji coba klasifikasi dengan algoritma *naive bayes classification* terhadap data set status huni rumah bantuan dana rehab rekon pasca erupsi Merapi 2010, dilakukan lebih dari satu kali sampai dengan mendapatkan hasil nilai akurasi tertinggi dan konstan/konsisten.

3.5 Rincian jalan/alur skema penelitian

Berikut di bawah ini merupakan gambar rincian alur skema penelitian yang dilakukan :



Gambar 1. rincian alur skema penelitian

4. HASIL DAN PEMBAHASAN

Uji Coba atau *Eksperimental* pada penelitian ini menggunakan validasi data *10-fold cross validation*, dimana *10-fold cross validation* memiliki nilai bias varian relatif rendah. Pada *10-fold cross validation*, data dibagi menjadi 10 bagian terlebih dahulu secara acak dengan perbandingan yang sama, lalu hitung *error rate* setiap bagian, setelah itu diperoleh *error rate* secara keseluruhan dari menghitung rata-rata *error rate* dari semua bagian data [4]. Untuk seluruh dataset yang digunakan pada penelitian ini adalah sebanyak 2516 *record/instances*. Pengujian dilakukan menggunakan tools pengolah data (framework) *RapidMiner*. Pengujian ini dilakukan lebih dari satu kali untuk memperoleh nilai akurasi klasifikasi terhadap data set status huni (sudah dihuni dan belum dihuni) dengan hasil yang paling tinggi dan konsisten. Adapun hasil uji coba/eksperimen ini dapat dilihat pada tabel Tabel 4 berikut ini :

Tabel 4 Hasil akurasi Algoritma *Naive Bayes Classification* untuk klasifikasi data status huni bantuan dana rumah rehabilitasi rekonstruksi pasca bencana erupsi Merapi 2010

Ekperimen Ke	Nilai Akurasi	Waktu eksekusi yang dibutuhkan
1	89,59 %	0.5 s
2	89,59 %	0.5 s
3	89,59 %	0.5 s

Hasil pada tabel 4 di atas, menunjukkan bahwa setelah dilakukan eksperimen sebanyak tiga kali, hasil nilai akurasi menunjukkan nilai stabil, terbaik/tertinggi dan konstan dari awal uji coba. Nilai akurasi klasifikasi data status huni (sudah dihuni dan belum dihuni) diperoleh nilai sebesar 89,59% dengan waktu eksekusi di bawah 1 detik.

Berikutnya hasil eksperimen ini dilanjutkan dengan menguji hasil nilai akurasi klasifikasi dengan menggunakan tabel *confusion matrix*. Hasil tabel *confusion matrix* dengan algoritma *Naive Bayes* ditunjukkan pada tabel 5 berikut seperti di bawah ini :

Tabel 5 *confusion matrix table* klasifikasi data status huni bantuan dana rumah rehab rekon pasca bencana erupsi Merapi 2010 dengan Algoritma *Naive Bayes*

	True Sudah	True Belum
Pred.Sudah	A 2053	b 169
Pred.Belum	C 93	d 201

Tabel *confusion matrix* di atas selanjutnya dilakukan pengujian dengan menghitung nilai akurasi berdasarkan persamaan (3). Hasil perhitungan :

$$Akurasi = \left(\frac{2053 + 201}{2053 + 169 + 93 + 201} \right) = \left(\frac{2254}{2516} \right) \times 100\% = 89,59\%$$

Berdasarkan hasil perhitungan akurasi secara manual seperti terlihat di atas, perhitungan persentase tingkat akurasi klasifikasi dengan algoritma *Naive Bayes Classification* terhadap data status huni (sudah dihuni dan belum dihuni) bantuan dana rumah rehab rekon pasca bencana erupsi Merapi 2010, dicapai nilai persentase sebesar 89,59%.

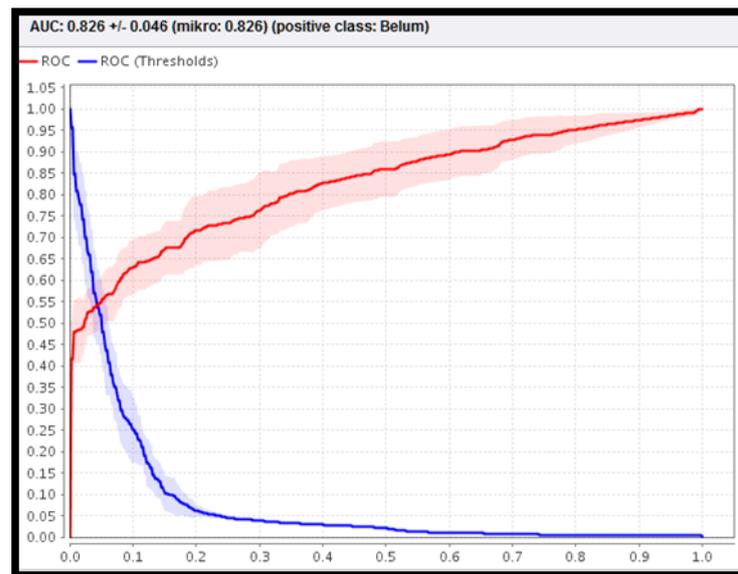
Pengujian performa hasil klasifikasi Algoritma *Naive Bayes* dilakukan dengan menggunakan *tools Rapidminer* dengan cara melihat hasil keluaran berupa kurva *Receiver Operating Characteristic curve (ROC)* dengan hasil nilai *Area Under the ROC Curve (AUC)* yang ditunjukkan pada tabel 5 berikut di bawah ini :

Tabel 5 Hasil akurasi dan *Area Under Curve (AUC)* klasifikasi data status huni bantuan dana rumah rehab rekon pasca bencana erupsi Merapi 2010 dengan Algoritma *Naive Bayes Classification*

Eksperimen Ke	<i>Naive Bayes Classification</i>		Waktu eksekusi yang dibutuhkan
	Nilai Akurasi	Performa AUC	
1	89,59 %	0.826	0.5 s
2	89,59 %	0.826	0.5 s
3	89,59 %	0.826	0.5 s

Tabel 5 di atas, dapat menunjukkan hasil bahwa eksperimen atau uji coba dilakukan sebanyak 3 (tiga) kali. Hal ini dilakukan untuk mendapatkan nilai akurasi dan performa AUC tertinggi serta konstan/konsisten. Dari percobaan pertama hingga ketiga nilai akurasi dan performa AUC menunjukkan hasil yang konsisten. Nilai Performa AUC mencapai 0.826. Hasil performa klasifikasi dengan nilai 0.826 tergolong ke dalam klasifikasi berperforma *good classification*. Nilai di antara 0.80 - 0.90 = klasifikasi baik atau *good classification* [7].

Visualisasi kurva AUC dalam bentuk gambar grafik dapat ditunjukkan pada gambar 2 berikut di bawah ini :



Gambar 2. Kurva ROC – AUC hasil eksperimen klasifikasi data status huni bantuan dana rumah rehab rekon pasca bencana erupsi Merapi 2010 dengan algoritma *Naive Bayes Classification*

5. KESIMPULAN

Berdasarkan hasil dan pembahasan yang telah disampaikan di atas, dapat disimpulkan bahwa algoritma klasifikasi Naive Bayes dapat diterapkan untuk melakukan klasifikasi data status huni bantuan dana rumah rehab rekon pasca bencana erupsi Merapi 2010 dengan hasil akurasi klasifikasi mencapai nilai sebesar 89,59%. Hasil performa akurasi klasifikasi diperoleh nilai AUC (*Area Under the ROC curve*) sebesar 0.826, maka Algoritma Naive Bayes dapat diterapkan untuk mengklasifikasi data status huni hunian rumah bantuan dana rumah (BDR) rehab rekon pasca erupsi Merapi 2010 dengan kategori baik.

Penelitian ini masih dapat dikembangkan dalam hal peningkatan hasil akurasi klasifikasi data status huni. Pengembangan penelitian ini dapat dilakukan dengan mengkombinasikan atau menambahkan metode algoritma fitur seleksi maupun algoritma optimasi bobot, misalnya : *Particle Swarm Optimization* (PSO), *Genetic Algorithm* (GA), *Chi Square*, *Ant Colony Optimization* (ACO), *Adaboost*. Besar harapan hasil penelitian ini dapat digunakan sebagai metode pengklasifikasi sekaligus prediksi berbasis probabilitas pada data status huni hunian rumah terbangun (hunian tetap) bagi para penerima bantuan dana rumah pasca bencana, baik pasca bencana gunung berapi, tsunami, banjir, gempa bumi, tanah longsor, kebakaran, angin puting beliung/tornado dan kebencanaan lainnya.

DAFTAR PUSTAKA

- [1] Beki, S., 2013. *Pendampingan yang mencerahkan*. Cetakan Pertama. Kementerian Pekerjaan Umum, Direktorat Jenderal Cipta Karya Republik Indonesia
- [2] KPI, 2014. *Key Performance Indicators : Community-based Settlements Rehabilitation and Reconstruction Project* (Rekompak), KPI Kompak JRF-PSF Status November 2014. tersedia di <http://www.rekompakciptakarya.org/KPI> [diakses : 18 Mei 2016].
- [3] Bramer, M., 2011. *Principles of data mining*, London, Springer

- [4] Han J., dan Kamber M., 2011. *Data Mining: Concepts, Models, and Techniques*, Verlag Berlin Heidelberg, Springer
- [5] Kusriani, dan Luthfi T.E., 2009. *Algoritma Data Mining*. P. Theresia Ari, Ed. Yogyakarta, Indonesia: Andi Offset.
- [6] Bustami., 2014 . *Penerapan Algoritma Naive Bayes untuk mengklasifikasi data nasabah asuransi*. Jurnal Informatika Vol. 8, No 1, hlm. 884-898.
- [7] Gorunescu, F., 2011. *Data Mining Concept, Models and Techniques, 12th ed., Prof. Lakhmi C. Jain Prof. Janusz Kacprzyk, Ed.* Craiova, Romania, Springer
- [8] Polczynski's D.R.L., Lecture. 2010. *WEKA Classification Using Decision Trees*. Computer science : college of enggining & Applied sciences.