



Studi Komparasi Metode SVM dan Naive Bayes pada Data Bencana Banjir di Indonesia

Riska K Abdullah¹, Ema Utami²

Fakultas Teknik, Program Studi Teknik Elektro Universitas Ichsan Gorontalo¹

Magister Teknik Informatika (MTI) Universitas AMIKOM Yogyakarta²

Email: riskakurnianto@unisan.ac.id¹, ema.u@amikom.ac.id²

Abstrak

Studi Komparasi Metode SVM dan *Naive Bayes* Pada Data Bencana Banjir di Indonesia bertujuan untuk mendapatkan *dataset clean* yang berisi bencana banjir lengkap dengan atribut cuaca. Pada *dataset* tersebut kemudian diimplementasikan model klasifikasi SVM dan *Naive Bayes*. Proses ini dilakukan agar performa antara SVM dan *Naive Bayes* dapat terlihat dan bisa dinilai mana yang lebih baik ketika diterapkan pada data bencana banjir di Indonesia. Penelitian dibagi menjadi tiga tahap utama, tahap pertama yaitu proses ekstraksi *dataset*. Proses tersebut bertujuan untuk mendapatkan *dataset* yang *clean*. Proses tersebut dilaksanakan dengan penerapan teknik data *mining* untuk menyatukan data cuaca dan data bencana alam berdasarkan tanggal dan lokasi kejadian. Tahap kedua yaitu proses implementasi klasifikasi, dan tahap terakhir yaitu proses *capturing* performa dari kedua model. Pada tahap terakhir pengukuran performa dari kedua model (SVM dan *Naive Bayes*) didapatkan dari Perhitungan akurasi dengan memanfaatkan *confusion matrix*, analisa ROC, kemudian parameter perbandingan selanjutnya yaitu waktu eksekusi. Hasil dari penelitian menunjukkan persentase Nilai akurasi rata-rata dari model SVM sebesar 48,90% sedangkan nilai akurasi dari *Naive Bayes* sebesar 64,70%. Sementara itu untuk masing-masing *runtime* SVM kurang lebih sebesar 720 mili detik dan *naive bayes* kurang lebih 280 mili detik. Dapat disimpulkan bahwa metode *Naive Bayes* lebih baik performanya dibandingkan dengan metode SVM ketika kedua metode tersebut diterapkan pada *dataset* yang sama yaitu *dataset* bencana banjir di Indonesia. Begitu pun dengan *runtime*, *Naive Bayes* masih lebih unggul karena memiliki waktu yang lebih singkat dalam proses *training* dan testing dibandingkan dengan SVM.

Kata kunci: *svm, naïve bayes, comparative, classification*

Abstract

Comparative Study of SVM and Naive Bayes Methods on Flood Disaster Data in Indonesia aims to obtain a clean dataset containing complete flood disasters with weather attributes. The dataset is then implemented by SVM and Naive Bayes classification model. This process have to be execute so the performance between SVM and Naive Bayes can be seen and can be assessed which is better when applied to flood disaster database in Indonesia. The research is divided into three main stages, the first stage is the extraction process dataset. The process aims to get a clean dataset. The process is carried out with the application of data mining techniques to unify weather data and natural disaster data based on date and location of incidents. The second stage is the process of classification implementation, and the last stage is the process of capturing the performance of both models. In the last stage the performance measurements of both models (SVM and Naive Bayes) are obtained from the calculation of accuracy by utilizing confusion matrix, ROC analysis, then the next comparison parameter is the execution time. The result of the research shows the percentage of average accuracy value from SVM model is 48,90% while accuracy value from Naive Bayes is 64,70%. Meanwhile for each SVM runtime approximately 720ms and Naive Bayes approximately 280 ms. It can be concluded that Naive Bayes method is better than the SVM method performance when both methods are applied to the same dataset that is the flood disaster dataset in Indonesia. So even with the runtime, Naive Bayes is still superior because it has a shorter time to execute the process of training and testing compared with SVM.

Keywords: *flood, svm, naïve bayes, comparative, classification*

A. PENDAHULUAN

Informasi tentang bencana alam menarik untuk dilihat dan dilakukan penelitian. Hal ini dimaksudkan agar bencana alam yang terjadi kembali dikemudian hari bisa lebih siap untuk dihadapi dengan persiapan yang matang tentunya. Kejadian bencana alam yang sering terjadi di Indonesia beberapa tahun terakhir ini saat penelitian ini ditulis yaitu bencana Banjir. Pola bencana banjir ini sangat menarik untuk ditinjau sehingga para pembaca ataupun peneliti bisa melihat pola yang tersembunyi di dalamnya. Untuk mewujudkan hal tersebut maka sebelumnya perlu

dilakukan proses ekstraksi *dataset* secara khusus dengan menggabungkan antara data bencana alam dan data cuaca di Indonesia. Dari kegiatan tersebut kemudian akan diterapkan teknik klasifikasi yaitu SVM dan *Naive Bayes*. Setelah itu, maka bisa didapatkan teknik klasifikasi mana yang cocok untuk diterapkan pada data cuaca dan bencana alam Indonesia.

Sementara itu tujuan dari perbandingan kedua metode klasifikasi antara SVM dan *Naive bayes* dimaksudkan agar ketika ada penelitian selanjutnya dibidang dan topik yang sama maka sudah ada jawaban teknik klasifikasi mana yang cocok untuk digunakan pada *database* bencana banjir di Indonesia. Kedua teknik ini diambil karena merupakan teknik yang sudah sangat familier yang diterapkan oleh para peneliti dalam melakukan proses klasifikasi.

Tinjauan Pustaka

Penelitian yang dilakukan oleh (Setianingsih & Hakim, 2015) yaitu menerapkan teknik asosiasi pada *database* bencana alam Indonesia yang ada di Indonesia pada *database* bencana tanah longsor. Pembahasannya didapat bahwa teknik asosiasi yang dilakukan bisa menghasilkan informasi yang menarik yaitu Jika terjadi tanah longsor dengan kondisi tanah labil maka hujan deras dengan nilai *support* sebesar 0,5227273 dan nilai *confident* sebesar 0,8214286. Dari hasil tersebut terlihat sedikit janggal karena adanya longsor sudah pasti disebabkan oleh hujan, bukan longsor menyebabkan hujan (Setianingsih & Hakim, 2015). Metode ekstraksi *dataset* pada penelitian tersebut akan diadopsi untuk diterapkan pada penelitian selanjutnya.

Pada penelitian (Susanti & Hidayanto, 2014) implementasi data *warehouse* spasial serta *dashboard* spasial terhadap kejadian bencana. Implementasi ini mempunyai dampak untuk mempermudah pemanfaatan informasi terkait bencana. Hasil penelitian ini memberikan rekomendasi berupa lokasi kejadian yang akurat, sehingga penanganan bencana dapat tepat dan sesuai sasaran. Kelebihan dari penelitian ini yaitu menerapkan data *warehouse* spasial untuk keperluan visualisasi posisi bencana alam (Susanti & Hidayanto, 2014).

Dalam hasil penelitiannya (Rupaka, Sudarno, & Suharyanto, 2013) didapati Kurangnya kesadaran pemerintah desa untuk memberikan pengarahan kepada warganya tentang bahaya longsor serta penyebabnya. Semestinya Badan Penanggulangan Bencana Daerah selaku bagian pemerintah yang menangani tentang masalah bencana alam harus meningkatkan kinerja, terutama dibidang pelatihan bahaya bencana alam kepada para pemangku kepentingan di desa atau kecamatan. Dari penelitian ini hasil penelitian belum terlalu baik dalam merepresentasikan keadaan karena hasil yang disimpulkan hanya berupa himbauan untuk masyarakat sama sekali tidak mengungkapkan hasil dari klasifikasi kerentanan yang dimaksud (Rupaka, Sudarno, & Suharyanto, 2013).

Manajemen bencana menghadapi masalah dalam menggunakan informasi terkini mengenai kebencanaan, untuk analisis terhadap situasi yang tepat dan pengambilan keputusan. Untuk mengatasi masalah tsb, (Basu, Bandyopadhyay, & Ghosh, 2016) mengusulkan dukungan *framework* yang mengumpulkan Informasi situasional melalui interaktif orang-orang menggunakan SMS dari "crowd" (kerumunan) yang hadir di lokasi bencana, dan merangkumnya berdasarkan tanggapan keadaan situasional sehingga bisa digunakan sebagai pengambilan keputusan

mengenai kerusakan. Sejak penggunaan *smartphone* di pedesaan India yang sangat rendah, maka fasilitas SMS dari "crowd" yang dipilih digunakan dalam membentuk informasi dari tempat penyimpanan inti. Tempat penyimpanan ini (*Repository*) ini digunakan untuk dukungan analisis situasi pasca-bencana. Metode yang digunakan dalam penelitian ini yaitu ekstraksi *text mining* dari *twitter* dan pesan singkat untuk mendapatkan informasi yang penting seperti contoh yaitu parameter ketersediaan dokter yang bisa disimpulkan dari pesan singkat yang berisi "Rumah sakit masih bisa menangani namun para masyarakat yang sedang dirawat tidak merasa nyaman." Teknik yang digunakan yaitu *automatic summarization* (Basu, Bandyopadhyay, & Ghosh, 2016).

Dalam penelitiannya (Matuska, Hudec, Kamencay, Benco, & Zachariasova, 2014) disajikan pengembangan metode baru untuk pengenalan obyek berdasarkan *hybrid local descriptors*. Metode ini memanfaatkan Kombinasi dari beberapa pendekatan (SIFT – *Scale Invariant Feature Transform*, SURF - *Speeded Up Robust Features*) dan terdiri dari dua bagian. Penerapan metode *hybrid* yang disajikan untuk beberapa gambar dari *dataset*. *Class* dari *dataset* mewakili hewan besar yang terletak di negara Slowakia, yaitu serigala, rubah, beruang coklat, rusa dan babi hutan. Metode yang disajikan dapat juga digunakan di daerah lain untuk klasifikasi citra dan ekstraksi ciri (*feature*). Hasil eksperimen menunjukkan, bahwa kombinasi *descriptor* lokal memiliki efek positif bagi pengenalan obyek. Metode dari penelitian ini yaitu : dilakukannya proses pengenalan objek, kemudian ekstraksi visual *descriptor* yang memanfaatkan *neighbour pixel*, lalu dilangkah terakhir dilakukan klasifikasi (Matuska, Hudec, Kamencay, Benco, & Zachariasova, 2014).

B. METODE

Pendekatan Penelitian

Obyek yang diteliti yaitu peristiwa bencana banjir dan keadaan cuaca yang ada di Indonesia. Pada obyek tersebut jika ditelusuri secara detail, setelah itu secara otomatis akan terlihat hal-hal mendasar apa saja yang menyebabkan bencana banjir terjadi. Hal tersebut dimungkinkan dengan melihat keadaan cuaca pada hari dan tanggal kejadian. Secara mendasar Penelitian ini berupa penelitian mandiri. Penelitian ini berkonsentrasi pada pengolahan data bencana alam dan data cuaca yang ada di Indonesia. Data diperoleh dari situs BNPB dan stasiun cuaca di Indonesia. Dengan detail data bencana alam dari BNPB dan data cuaca dari stasiun cuaca yang ada di Indonesia. Selanjutnya dari kedua data tersebut akan diterapkan komparasi metode untuk klasifikasi. Klasifikasi data *mining* yang digunakan ialah metode *naive bayes* dan *svm*. Kedua metode ini dibandingkan satu sama lain dan *goal* teknik klasifikasi mana yang lebih cocok bila digunakan pada data bencana alam dan data cuaca. Parameter yang digunakan untuk mengukur perbandingan dari kedua metode ini yang yaitu : akurasi, analisa ROC dan AUC.

Metode Pengumpulan Data

Adapun pengumpulan data dilakukan dengan mengambil langsung dari situs instansi yang terkait. Untuk data kejadian banjir didapatkan dari situs BPPB Indonesia. Data selanjutnya yaitu data informasi cuaca yang memuat atribut suhu rata-rata harian, kecepatan angin dan curah hujan didapat dari instansi NOAA-NCEI (National Oceanic And Atmospheric Administration - National Centers for Environmental Information).

Sementara itu data ketiga yaitu data posisi dan nama stasiun cuaca yang ada di Indonesia didapatkan dari situs instansi NOAA-NCEI juga.

Metode Analisis Data

Analisis data dilakukan dengan menggunakan *tools* bahasa pemrograman *python* dan beberapa perangkat lunak penunjang lainnya. Untuk keperluan analisis ini akan diterapkan konsep dan metode klasifikasi SVM dan *Naïve bayes*. Setelah diterapkannya implementasi klasifikasi ke dalam bentuk program maka selanjutnya dianalisis parameter yang dihasilkan oleh masing-masing program dari tingkat akurasi serta kecepatan waktu yang dibutuhkan untuk memproses data *training* dan data tes. Dari masing-masing program tersebut diambil nilai akurasi saat diterapkan pada *database* bencana banjir. Sementara itu untuk membandingkan hasil dari klasifikasi maka dilakukan uji dengan program standar bawaan yaitu dengan perangkat lunak *data mining* Rapid Miner.

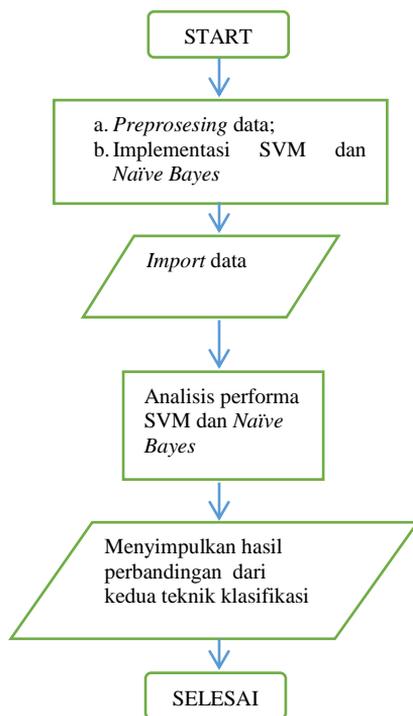
Alur Penelitian

Adapun langkah – langkah penelitian ini dilakukan secara bertahap sesuai dengan alur di bawah ini :

1. *Preprocessing / wrangling* data
 - a. Ekstraksi data bencana banjir dengan jarak stasiun cuaca
 - b. Ekstraksi data bencana banjir dengan nama stasiun yang mempunyai jarak minum
 - c. Ekstraksi *dataset* bencana banjir lengkap disertai atribut cuaca saat kejadian.
2. Implementasi Teknik SVM dan *Naïve Bayes*.

3. *Debugging* untuk meminimalkan *error* yang terjadi
4. Mengukur performa model klasifikasi dengan ROC, *Confusion Matrix* dan jumlah waktu eksekusi.
5. Membandingkan hasil yang dieksekusi dengan program standar pengolahan *datamining (Rapidminer)*
6. Menarik kesimpulan hasil penelitian mengenai metode yang tepat untuk *database* bencana banjir di Indonesia.

Proses alur penelitian dapat dilihat pada gambar 1. *flowchart* penelitian.



Gambar 1. Alur Penelitian

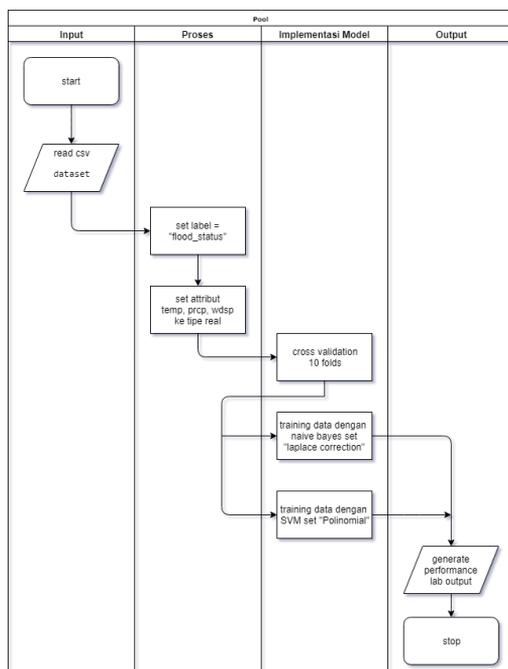
C. HASIL DAN PEMBAHASAN

Rancangan Sistem

Untuk rancangan sistem model klasifikasi terdapat pada gambar 2. Pada awal rancangan direncanakan masukan disimpan ke dalam *dataframe* dengan nama variabel dataset. Variabel ini berasal dari berkas berformat *csv*. Perlu diketahui berkas ini diperoleh dari tahap proses ekstraksi *dataset* tahap ketiga sebelumnya. Setelah variabel dimuat ke dalam lingkungan pengembangan (*development environment*), kolom *flood_status* di beri *flag* label. Hal ini dimaksudkan agar proses *training* klasifikasi keluarannya tertuju pada kolom *flood_status*. Sementara itu atribut sisanya yaitu atribut suhu, curah hujan dan kecepatan angin yang diberi nama berurutan *temp*, *prcp* dan *wdsp*. Atribut ini yang akan menjadi *feature* dari klasifikasi.

Khusus untuk *training* masukan sampel klasifikasi SVM dan *Naive Bayes* sampel digunakan metode *cross validation* dengan parameter 10 *folds*. Sementara itu untuk parameter SVM digunakan metode *polynomial*. Pemilihan metode tersebut didasarkan dari keluaran *flood_status* yang bersifat *binominal* yaitu berupa keluaran yang bersifat *categorical* dalam kasus ini keluarannya spesifik hanya bernilai dua kondisi yaitu banjir dan tidak banjir. Untuk mendekati hal tersebut maka digunakan metode *polynomial* pada klasifikasi SVM.

Setelah data *training* digunakan maka selanjutnya parameter dari hasil klasifikasi berupa kecepatan eksekusi, nilai akurasi, dan *recall* akan *digenerate* dalam bentuk HTML berupa *report*.



Gambar 2. Rancangan Sistem

Preprocessing Data

Tahap *preprocessing* data merupakan tahap yang sangat panjang dalam proses ini. Pada tahap ini data mentah yang ada, yaitu data bencana banjir dan data cuaca akan digabung menjadi satu dengan mengawinkan parameter lokasi. Data atribut untuk bencana banjir dapat dilihat pada tabel 1 sedangkan atribut untuk data cuaca dapat dilihat pada tabel 3. Jumlah baris pada Tabel 2 Sebanyak 1309 baris, kemudian jumlah baris dari Tabel 3 sebanyak 100753 baris.

Tabel 1 Penggalan Data Kejadian Banjir

Bujur	Lintang	Lokasi	Korban	Kerugian	Keterangan
124,815	0,988698	Kec.Pusamen,Kab. Minahasa Tenggara, Sulut	Nihil	Rumah : 55 RB, 587 RR, 20 hanyut	

Bujur	Lintang	Lokasi	Korban	Kerugian	Keterangan
112,532	-177,712	Ds. Tb. Sanggai Ds. Tb. Mangkup Ds. Rantau Kantang Ds. Tukang Langit Ds. Tb. Boloji Ds. Bajane Ds. Tb. Kowan Ds. Rantau Tampang Kec. Telaga Antang Kab. Kotawaringin Timur Prov. Kalimantan Tengah	125 KK terdampak		
975,483	2,977	Ds. Jambo Keupok, Ds. Beutong, Ds. Ujong Gunong Cut, Ds. Ujong Gunong Rayeuk, Ds. Rambong, Ds. Buket Gandeng, Ds. Seunubok Keuranji, Ds. Ujong Tanoh, Ds. Alor Duamas, Kec. Kota Bahagia, Kab. Aceh Selatan, Prov. Aceh.	Nihil		Kronologis : Hujan lebat selama 2 hari Upaya : BPBD Aceh selatan telah mengirimkan logistic. BPBD Aceh selatan < Dinsos, TNI dan Polri tengah berada dilokasi kejadian untuk melakukan assesment

Kedua data ini yaitu pada tabel 1 dan tabel 2 kemudian digabungkan menjadi satu dengan mencocokkan parameter lokasi bencana dari data bencana banjir dan lokasi stasiun cuaca dari data cuaca. Lokasi stasiun cuaca dan koordinatnya contoh tabelnya dapat dilihat pada tabel 3. Jumlah stasiun cuaca sampai saat penelitian ini ditulis yaitu sebanyak 164 stasiun yang tersebar di seluruh provinsi yang ada di Indonesia.

Tabel 2 Penggalan data cuaca tahun 2011 – 2017

STN---	WBAN	YEARMODA	TEMP	WDSP	PRCP	SNDP
960010	99999	20110101	80,8	5,2	0,00I	999,9
960010	99999	20110102	81,4	4,8	0,00I	999,9
...
979800	99999	20170331	81,6	3,4	99,99	999,9
979800	99999	20170401	82,8	3,1	0,31A	999,9

Berdasarkan data dari Tabel 2, Tabel 3 dan Tabel 4 maka langkah selanjutnya yaitu mencari jarak stasiun cuaca dengan lokasi bencana. Hasil dari perhitungan ini dapat dilihat pada tabel 5.

Perhitungan jarak antara lokasi bencana dan stasiun cuaca dihitung menggunakan metode *Vincenty*. Metode ini cocok digunakan karena perhitungan jarak antara satu titik dengan titik lainnya tetap memperhitungkan nilai kelengkungan permukaan bumi. Lebih jelas mengenai teori ini dapat dilihat pada bab 2. Perhitungan dilakukan dengan membandingkan kurang lebih 164 stasiun cuaca dengan masing-masing lokasi bencana dan menghitung jaraknya. Setelah jarak antar lokasi bencana dan masing-masing stasiun cuaca didapatkan, maka langkah selanjutnya yaitu mendapatkan jarak terdekat lokasi stasiun cuaca dari lokasi bencana banjir.

Tabel 3. Penggalan Data Stasiun Cuaca

USAF	STATION NAME	LAT	LON
749308	PULAU FANI	1083	131,267
749402	KAMERI	-0,95	134,817
749403	SANSAPOR	-0,5	132,083
...
749539	SERUMI	-1,233	136,2
960010	MAIMUN SALEH	5,874	95,34
960090	MALIKUS SALEH	5,227	96,95
960110	SULTAN ISKANDARMUDA	5,524	95,42

Setelah jarak antar setiap stasiun dan lokasi bencana didapatkan maka langkah selanjutnya yaitu memasukkan *flag* status terjadinya bencana pada data cuaca berdasarkan lokasi stasiun cuaca dan tanggal kejadian bencana. Hasil dari proses ini dapat dilihat pada Tabel 5. Adapun atribut yang diambil sebagai bahan acuan dari penelitian ini yaitu atribut, temperatur rata-rata harian (*temp* dengan satuan *celcius*), kecepatan rata-rata angin harian (*wdsp* dengan satuan *knots*) dan curah hujan rata-rata harian (*prcp* dengan satuan cm^2). Pada Tabel 5 jumlah baris *dataset* berjumlah 564 Baris.

Tabel 5. Penggalan *dataset* cuaca dengan status banjir

ind	temp	prcp	flood_status	wdsp
0	22,72	80,52	1	10,4
1	25,61	2,58	1	3,4
2	28,61	1,03	1	3,6
3	26,83	5,1	1	1,7
4	26	3,68	1	1,9
...
559	27,33	0,19	0	3,5
560	27,56	0,52	0	1,4
561	27,17	0,52	0	3,2
562	26,11	0,77	0	2,6
563	26,11	1,03	0	3,7

Implementasi Model Klasifikasi

Model yang diimplementasikan untuk percobaan secara detail menggunakan *tools* pada Tabel 6. Khusus untuk implementasi *Naive Bayes* parameter yang di masukkan yaitu model *naive bayes* menggunakan parameter *laplace correction*.

Tabel 6. Spesifikasi tools yang digunakan

Tools	Tipe	Versi
Orange	Perangkat lunak Data Mining	13.0
Rapid Miner	Perangkat lunak Data Mining	8.2.001
Numpy	<i>Library python</i>	1.12.1
Pandas	<i>Library python</i>	0.19.2
Geopy	<i>Library python</i>	1.11.0
Jupyter Notebook	<i>Library python</i>	1.0.0
Python	Bahasa pemrograman	3.5

Untuk implementasi model SVM diberikan parameter tipe *kernel*: *Polinomial*. Jumlah iterasi sebanyak 100000 dan *kernel cache* sebesar 200. Parameter ini dipilih karena merupakan parameter standar bawaan yang ada pada perangkat lunak *data mining orange*.

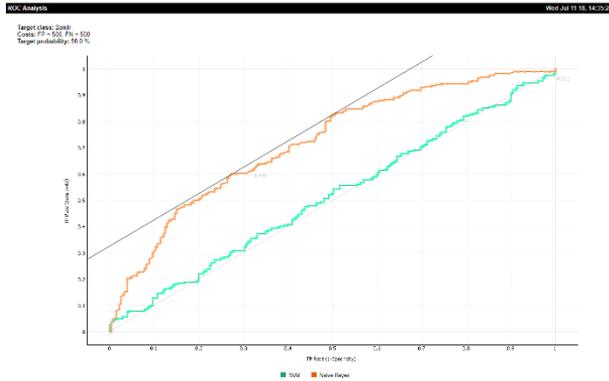
Analisis Hasil

Analisis hasil dari implementasi dan pengukuran performa model menggunakan perbandingan ROC, *Confusion Matrix*, *Lift Curve*. Sementara itu untuk melihat implementasi dari model tersebut apakah *reliabel* atau tidak implementasi model dibandingkan dengan perangkat lunak lain yang hasil analisisnya ada di sub bab perbandingan model dengan standar lain. Secara detail masing-masing analisis sebagai berikut.

Analisis ROC

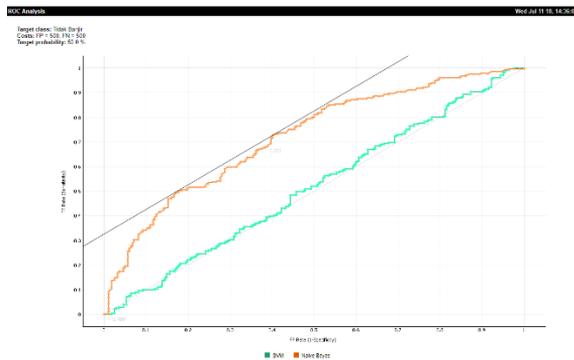
Pada Gambar 3 dan Gambar 4 masing-masing merepresentasikan bentuk kurva ROC dengan target *class* Banjir dan Tidak Banjir secara berurutan. Secara sederhana model yang mempunyai luas lebih besar dari model lainnya merupakan model yang mempunyai tingkat akurasi lebih tinggi. *Sensitivity* adalah proporsi nilai yang diprediksi dengan benar di antara semua kemungkinan nilai yang benar. Sedangkan *specificity* adalah

proporsi dari nilai yang benar di antara semua nilai yang berhasil diprediksi.



Gambar 3. Kurva Perbandingan ROC dengan target class banjir

Secara sederhana dapat dilihat *Naive Bayes* memiliki luas area di bawah kurva yang lebih besar dibandingkan dengan SVM baik pada target *class* banjir dan tidak banjir. Hal tersebut menunjukkan tingkat keakuratan dari *Naive Bayes* lebih baik daripada SVM.



Gambar 4. Kurva ROC dengan target class tidak banjir

Confusion Matrix

Confusion Matrix akan menjadi dasar dari perhitungan akurasi dari suatu model. Pada gambar 5 nilai TP = 2 , FN = 280, FP = 8, dan TN

= 274. Dengan menggunakan rumus (2) dan rumus (3). *Error rate* didapatkan sebesar = 0,512 maka akurasi = 0,4884 atau sebesar 48,84%.

		Predicted		
		Tidak Banjir	Banjir	Σ
Actual	Tidak Banjir	20.0 %	50.5 %	282
	Banjir	80.0 %	49.5 %	282
Σ		10	554	564

Gambar 5. Confusion matrix untuk model SVM

Pada Gambar 6 ditunjukkan nilai TP = 193, TN = 172, FP = 110, FN = 89. Dengan menggunakan rumus yang sama yaitu rumus (1) dan (2) maka nilai *error rate* dari model *Naive Bayes* sebesar 0,3528. Dari nilai *error rate* tersebut didapat nilai akurasi SVM sebesar = 0,6472 atau 64,72%.

		Predicted		
		Tidak Banjir	Banjir	Σ
Actual	Tidak Banjir	63.7 %	34.1 %	282
	Banjir	36.3 %	65.9 %	282
Σ		303	261	564

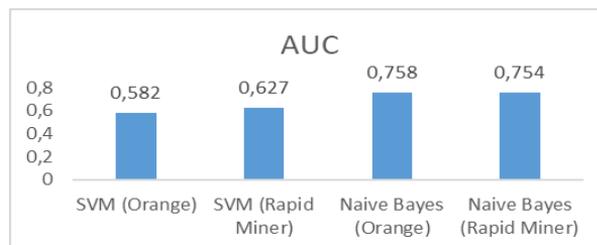
Gambar 6. Confusion matrix untuk model naive bayes

Tuliskan hasil penelitian dengan sekuens yang logis, sesuai dengan alur penelitian. Pada umumnya hasil diawali dengan karakteristik subjek penelitian. Tabel dan gambar disajikan secara informatif dengan jumlah maksimal 3-5 tabel atau gambar setiap naskah. Tabel harus *self explanatory*; angka-angka dan satuan harus tergambar dengan jelas pada tabel sehingga tidak tergantung pada teks. Judul tabel ditempatkan di atas tabel. Setiap tabel/gambar harus dinyatakan dalam naskah (teks).

D. PENUTUP

Simpulan dan Saran

Telah berhasil dilakukan ekstraksi *dataset* bencana banjir dari data cuaca di Indonesia dari tahun 2011 hingga 2017, data kejadian banjir, dan data posisi stasiun cuaca yang ada di Indonesia. Setelah ekstraksi *dataset* dilakukan telah berhasil juga diimplementasikan dan dilakukan uji coba model klasifikasi untuk studi perbandingan model klasifikasi SVM dan *Naive bayes*. Berdasarkan hasil implementasi dan percobaan, dapat diambil kesimpulan bahwa metode *Naive Bayes* lebih cocok untuk digunakan pada *dataset* untuk bencana banjir di Indonesia. Dengan nilai akurasi sebesar 64,7%, *naive bayes* mengungguli model SVM yang hanya mempunyai tingkat akurasi sebesar 48,9%. Sementara itu jika dilihat dari waktu eksekusi (*runtime*) *Naive Bayes* memiliki waktu yang sangat singkat untuk melakukan *training* dan juga tes. Jumlah waktu yang dibutuhkan oleh *Naive Bayes* hanya sebesar 280 mili detik, sedangkan SVM memerlukan waktu sebesar 720 mili detik.



Gambar 7. Perbandingan hasil AUC dengan Rapidminer

Dalam hal akurasi *Naive Bayes* tetap mengungguli SVM dengan perbedaan sekitar 15,8% lebih tinggi dari SVM. Sementara itu untuk perbedaan waktu eksekusi antara *Naive Bayes* dan SVM, SVM agak lambat sekitar 38% dibandingkan dengan *Naive Bayes*. Berdasarkan *dataset* yang terlibat pada kedua metode ini terlihat SVM yang sangat

populer menurut sciencedirect.com masih kurang cocok untuk digunakan jika dibandingkan dengan *Naive Bayes*. Namun dengan begitu metode SVM bisa saja diperbaiki dengan memasang CNN (*Convolutional Neural Network*) pada bagian awal sistem.

Dataset yang berhasil didapatkan pada penelitian ini selanjutnya dapat digunakan untuk percobaan klasifikasi yang bersifat *native* pada perangkat *mobile* ataupun *handled*. Perlu adanya percobaan tersebut sehingga bisa ditentukan *optimasi* seperti apa yang perlu dilakukan ketika model diimplementasikan dengan perangkat IoT. Setelah itu dari keseluruhan sistem jika digabung dapat menjadi satu kesatuan yaitu *Early Warning System* untuk bencana banjir.

Dataset yang diciptakan dalam penelitian ini diharapkan bisa digunakan oleh peneliti selanjutnya. *Dataset* tersebut dapat diunduh pada link berikut dengan nama berkas “dataset_banjir_based_on_tempr_prpc_v3.csv” (https://github.com/kurniantoska/id_flood_disaster_class). Tidak sebatas itu saja, namun seluruh kode program dari penelitian ini dapat ditemukan pada *link* tersebut. Kode program tersebut antara lain yaitu kode program yang ditulis untuk mengekstraksi *dataset* dari data kejadian banjir, data cuaca, dan data dari lokasi stasiun cuaca. Tidak lupa kiranya mohon untuk menyertakan pengutipan ataupun *copyright* ketika menggunakan *resouce* yang berkaitan dengan penelitian ini.

DAFTAR PUSTAKA

Basu, M., Bandyopadhyay, S., & Ghosh, S. (2016). Post disaster situation awareness and decision support through interactive crowdsourcing. *Procedia Engineering*, 159(June), 167–173. <https://doi.org/10.1016/j.proeng.2016.08.151>

- Matuska, S., Hudec, R., Kamencay, P., Benco, M., & Zachariasova, M. (2014). Classification of Wild Animals based on SVM and Local Descriptors. *AASRI Procedia*, 9(Csp), 25–30. <https://doi.org/10.1016/j.aasri.2014.09.006>
- Rupaka, A. P. G., Sudarno, & Suharyanto. (2013). Penilaian Potensi Bencana Longsor Berdasarkan Tingkat Kerentanan di Kabupaten Tegal. In S. P. Hadi, P. Purwanto, H. Rya Sunoko, & H. Purnaweni (Eds.), *Seminar Nasional Pengelolaan Sumberdaya Alam dan Lingkungan* (pp. 249–252). Semarang: Program Studi Ilmu Lingkungan - Program Pasca Sarjana Universitas Diponegoro, Semarang, Indonesia Jl. Imam Bardjo, SH No. 5 Semarang 50241 Telp/Fax. (024)8453635, 8452770.
- Setianingsih, D., & Hakim, R. F. (2015). Penerapan data mining dalam analisis kejadian tanah longsor di indonesia dengan menggunakan association rule algoritma apriori. In *Prosiding Seminar Nasional Matematika dan Pendidikan Matematika UMS 2015* (pp. 731–741). Indonesia.
- Susanti, I., & Hidayanto, A. N. (2014). Perancangan dan Implementasi Data Warehouse Spasial untuk mendukung Layanan Kebencanaan : Studi ... Perancangan dan Implementasi Data Warehouse Spasial untuk mendukung. Jakarta. Retrieved from <https://www.researchgate.net/publication/265171608%0APerancangan>
- Turkington, T., Breinl, K., Ettema, J., Alkema, D., & Jetten, V. (2016). A new flood type classification method for use in climate change impact studies. *Weather and Climate Extremes*, 14(September), 1–16. <https://doi.org/10.1016/j.wace.2016.10.001>
- Ya'acob, N., Ismail, N. S., Mustafa, N., & Yusof, A. L. (2014). Investigation of flood pattern using ANOVA statistic and remote sensing in Malaysia. *IOP Conference Series: Earth and Environmental Science*, 20(1). <https://doi.org/10.1088/1755-1315/20/1/012030>

