

Penerapan Deteksi Bencana Banjir Menggunakan Metode Machine Learning

¹Nenden Siti Fatonah

Fakultas Ilmu Komputer, Teknik Informatika, Universitas Mercu Buana¹
Jl. Meruya Selatan, Kec. Kembangan, Jakarta Barat, Daerah Khusus Ibukota Jakarta 11650¹
Nendenfatolah@gmail.com¹

Abstract

Abstract - Natural disasters can be defined as an unexpected event, this often happens in Indonesia, especially floods that occur within 5 years, but lately the flood disaster has been felt almost every year in several parts of Indonesia. . Flooding can be interpreted as an event that appears suddenly quite quickly where the water discharge cannot be accommodated in the drainage channel (river trough) or the flow of water discharge that goes to the sea in the sewer is blocked, so that the area that is obstructed by the water discharge results in a flood. overflow of water discharge in reservoir channels in several surrounding areas and is one of the natural disasters that occur at an unexpected time that cannot be prevented, because of this it is necessary to strive for an application to detect how to cope with flood disasters for the following years to reduce the impact the loss it causes to the human population. In this study, how to detect flood disasters that have occurred in the previous month so that this year and the following year the parties involved and the community around the area can find out how big the impact of the flood disaster will be. This research uses the Data Mining method with the algorithm used is the Naïve Bayes algorithm. The results obtained in this study were 76.73% using the Naïve Bayes algorithm for its accuracy. However, the tools used have not been able to predict the day of the flood disaster.

Keyword: Natural Disaster, Flood Disaster, Data Mining, Naive Bayes.

Abstrak

Abstrak – Bencana alam dapat di definisikan sebagai sebuah kejadian yang tidak terduga terjadi nya, hal tersebut sering terjadi di Indonesia terutama bencana banjir yang terjadi dalam kurun waktu 5 tahunan namun akhir-akhir ini bencana banjir tersebut sudah dapat dirasakan hampir setiap tahun di beberapa bagian wilayah di indonesia. Banjir dapat diartikan sebagai suatu kejadian yang hadir dengan tiba-tiba secara cukup cepat di mana tidak tertampungnya debit air dalam saluran pembuangan (palung sungai) atau terhambatnya aliran debit air yang menuju ke laut di dalam saluran pembuangan, sehingga daerah yang terhambat dengan debit air mengakibatkan meluapnya debit air pada saluran penampungan pada beberapa daerah sekitarnya dan merupakan salah satu bencana alam yang terjadi dengan waktu yang tak terduga tidak dapat dicegah, karena hal tersebut maka perlu diupayakan sebuah penerapan untuk mendeteksi bagaimana dapat menanggulangi bencana banjir untuk tahun-tahun selanjutnya untuk mengurangi dampak kerugian yang diakibatkannya bagi populasi manusia. Pada penelitian ini dilakukan bagaimana dapat mendeteksi bencana banjir yang telah terjadi di bulan sebelumnya agar pada tahun ini dan tahun selanjutnya pihak yang terkait dan masyarakat sekitar daerah tersebut dapat mengetahui seberapa besarnya dampak bencana banjir yang akan terjadi. Penelitian kali ini menggunakan metode Data Mining dengan algoritma yang digunakan adalah algoritma Naïve Bayes. Hasil yang di hasilkan dalam penelitian ini adalah 76,73% dengan menggunakan algoritma Naïve Bayes untuk ke accuracyan nya. Namun tools yang digunakan belum mampu memprediksi hari terjadinya bencana banjir.

Keyword: Bencana Alam, Bencana Banjir, Data Mining, Naive Bayes.

I. PENDAHULUAN

Indonesia merupakan salah satu negara di wilayah asia tenggara yang memiliki relief struktur alam dan kekayaan sumber daya alam yang cukup melimpah dan sangat berkah tanah yang di miliki oleh indonesia. Sumber daya yang dimiliki oleh indonesia sangat melimpah baik sumber daya alam hayati maupun sumber daya alam non hayati. Kekayaan alam yang dapat di ambil potensinya di Indonesia terdiri dari kekayaan laut, darat, dan kekayaan alam lainnya yang terkandung di dalam perut bumi. Indonesia sebagai negara kepulauan secara geografis terletak di khatulistiwa, di antara Benua Asia dan Australia serta di antara Samudra Pasifik dan Hindia, berada pada pertemuan tiga lempeng tektonik utama dunia, mengakibatkan Indonesia sebagai wilayah teritorial yang sangat rawan terhadap beberapa bencana alam yang sering ada pada wilayah Indonesia. Pada saat kondisi iklim yang cukup tidak menentu seperti saat ini hal tersebut berpengaruh terhadap iklim di Indonesia, maka perubahan musim dapat menjadi sebuah pemicu terjadinya bencana banjir yang berkepanjangan, kekeringan yang berkepanjangan, dan kebakaran hutan baik secara alami atau buatan terjadinya yang lebih meluas lagi dari tahun-tahun sebelumnya. Banjir merupakan salah satu cerita yang sering didengar baik di kalangan kabar burung atau berita-berita online yang ada saat, dimana hal tersebut terjadi pada musim hujan dengan intensitas debit air yang sering, lebat, dan dalam jangka waktu yang relatif panjang, daerah yang menjadi langganan banjir pada musim penghujan ada disekitar kawasan arus sungai yang aktif mengalir dari hulu ke hilir atau bendungan yang kedalaman nya sudah cukup berkurang di akibatkan dari kelalaian manusia untuk taat dalam membuang sampah[1]. Banjir juga dapat didefinisikan sebagai wilayah yang memiliki debit air cukup tinggi melebihi dari kapasitas daya tampung bendungan atau

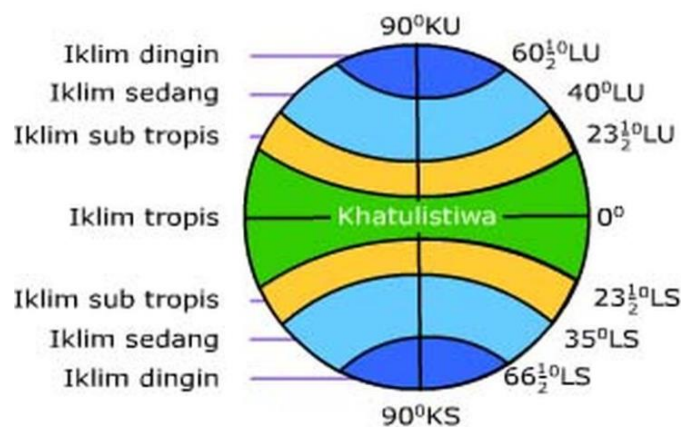
sungai pada daerah tersebut, hal tersebut dapat menimbulkan kerugian yang cukup tinggi karena menghancurkan harapan hidup masyarakat dengan menghilangkan sebagian atau semua kekayaan yang dimiliki baik yang berbentuk benda hidup, seperti anggota keluarga, ternak dan tanaman yang telah di tanam beberapa bulan untuk dapat di jual, seperti rumah, pekarangan, ladang, dan sawah tempat masyarakat menggantungkan hidup, fisik, sosial dan ekonomi[2]. Terjadinya bencana banjir yang terjadi hampir di setiap masuknya musim penghujan dapat juga dikarenakan oleh rendahnya kemampuan filtrasi tanah terhadap debit air yang cukup banyak namun tanah tersebut tidak cukup untuk menampung debit air yang cukup banyak jumlah debit air yang datang pada sebuah daerah dengan begitu tiba-tiba. Banjir dapat terjadi juga akibat meningkatnya permukaan air akibat dari curah hujan yang terjadi pada ambang diatas normal, perubahan suhu yang cukup signifikan dengan begitu cepatnya, sungai/bendungan yang bobol dikarenakan debit air yang cukup banyak, terhambatnya debit air di tempat lain dikarenakan dengan menumpuknya sampah-sampah yang di buang oleh masyarakat atau faktor alam yang pohon-pohon banyak berguguran yang jatuh pada aliran sungai yang debit airnya mengalami peningkatan[3].

Terjadinya bencana banjir dapat di jabarkan dengan beberapa faktor, baik faktor yang terjadi karena siklus alam maupun karena faktor kesalahan yang di lakukan oleh manusia terhadap alam sekitarnya, Berikut beberapa penyebab terjadinya banjir yang terjadi saat sekarang ini:

1. Hujan, Hujan yang terjadi pada saat ini seringnya terjadi dengan cukup lebat dan melebihi kapasitas sistem dari drainase yang tersedia di setiap wilayah di Indonesia. Terkadang hujan lebat terjadi dalam waktu relatif singkat karena jumlah hujan yang turun cukup tinggi dan di ikutin dengan angin dengan kecepatan knot cukup cepat yang dapat menyebabkan banjir di setiap wilayah. Di waktu yang berbeda, hujan gerimis selama berhari-hari atau berminggu-minggu dapat mengakibatkan banjir juga dikarenakan waktu hujan yang cukup panjang.
2. Sungai meluap, Sungai yang meluap dapat menyebabkan banjir dikarenakan sungai tersebut tidak dapat menampung jumlah debit air, banyak faktor yang mendukung kenapa sungai tersebut dapat meluap, salah satu dari akibat meluapnya sebuah sungai karena banyak sampah yang berserakan di sungai akibat nya aliran air tidak dapat mengalir dengan cepat agar dapat sampai ke laut untuk pembuangannya. Meluapnya sungai dapat terjadi ketika air lebih banyak di bagian hulu dari biasanya. Saat air itu mengalir ke hilir menuju dataran rendah yang berdekatan akan ada semburan dan air masuk ke daratan tanpa ada yang menghalangi debit air tersebut tumpah ke daratan.
3. Badai dan angin kencang di pantai dapat membawa angin kencang dan angin topan ke daratan pantai yang kering lalu menyebabkan banjir terjadi dengan begitu cepatnya. Kondisi tersebut dapat lebih memperburuk jika angin tersebut juga membawa hujan sendiri. Terkadang air dari laut akibat tsunami dapat mengalir ke daratan dan menyebabkan kerusakan pada daerah yang di lewati.
4. Bendungan rusak, Tanggul yang dibangun pada sisi sungai dan digunakan untuk mencegah debit air yang meningkat untuk membanjiri daratan yang berbatasan dengan sungai tersebut. Definisi Bendungan merupakan wadah dalam ukuran besar yang dibuat secara buatan atau bangunan yang kokoh yang dirancang untuk menahan laju debit air dari waduk, danau, atau tempat rekreasi untuk mencegah bencana banjir yang cukup besar pada sebuah daerah. Tetapi bila terlalu banyak air yang tertahan di bendungan dapat menyebabkan bendungan rusak dan air meluap ke daerah-daerah sekitarnya yang tidak memiliki bendungan. Kelebihan air juga bisa secara sengaja dialirkan dari bendungan untuk mencegah rusaknya bendungan yang berakibat banjir.
5. Es dan salju mencair di wilayah yang dingin yang mengakibatkan jumlah air di muka bumi meningkat, salju tebal selama musim dingin biasanya tetap tidak meleleh selama kurun waktu dari pembagian musim dari setiap wilayah yang ada di dunia. Ada juga gunung yang memiliki es di puncak gunung di beberapa wilayah yang ada di seluruh dunia. Terkadang es tersebut tiba-tiba meleleh ketika suhu bumi meningkat, mengakibatkan pergerakan besar debit air ke tempat-tempat yang ritme daerahnya sering kering menjadi meningkatnya ritme debit air. Pergerakan besar air ini biasanya disebut banjir pencairan salju (snowmelt flood).
6. Penyumbatan drainase, Banjir dapat terjadi ketika pencairan salju atau limpasan hujan tidak dapat disalurkan dengan tepat ke sistem drainase yang baik mengakibatkan air mengalir ke daratan yang tidak memiliki daerah serapan yang baik. Penyumbatan drainase atau kurang tepatnya sistem drainase biasanya menjadi penyebab banjir jenis ini. Daerah yang terjadi penyumbatan drainase akan tetap banjir sampai sistem air hujan atau saluran air diperbaiki oleh pemerintah terkait, drainase agar tetap baik juga butuh peran aktif masyarakat untuk dapat menjaga drainase tersebut tetap dapat berfungsi dengan seharusnya.

Definisi dari iklim adalah sebuah unsur geografi yang penting untuk melihat keadaan pada cuaca yang ada pada daerah tersebut, karena iklim memiliki pengaruh yang besar terhadap aktivitas manusia di lingkungan, seperti bidang transportasi, pertanian, perkebunan, dan lainnya. Climate atau arti dari bahasa indonesianya yaitu iklim adalah sintesis atau kesimpulan dari perubahan nilai unsur- unsur cuaca (hari demi hari dan bulan demi bulan sampai tahun demi tahun) dalam jangka waktu yang cukup panjang di suatu daerah atau pada suatu wilayah terdapat beberapa macam iklim yang tersebar pada seluruh dunia. Definisi Iklim lainnya dapat di artikan sebagai cuaca yang terjadi pada rentang waktu yang cukup panjang dan luas wilayah yang cukup luas. Cuaca memiliki definisi dimana kondisi udara pada wilayah di sekitaran wilayah tersebut dan dalam waktu yang relatif pendek. Cuaca dapat berubah dalam waktu yang cukup cepat dari waktu ke waktu yang ditandai dengan terjadinya siang dan malam pada setiap harinya. Salah satu faktor yang cukup signifikan dalam perubahan cuaca adalah curah hujan yang tinggi kadarnya. Indonesia memiliki tata letak di wilayah

daerah yang beriklim tropis yang menjadikan Indonesia memiliki curah hujan tinggi dan menerima pantulan cahaya yang cukup terik dari pencahayaan matahari yang sering terjadi pada setiap tahunnya dalam kurun waktu enam bulanan. Kejadian tersebut juga disertai tingginya resiko bencana akibat peristiwa iklim ekstrem seperti terjadinya banjir dan kemarau yang berkepanjangan. Peningkatan suhu global dunia diperkirakan akan menyebabkan perubahan-perubahan lainnya seperti meningkatnya permukaan air laut, meningkatnya intensitas fenomena cuaca ekstrem dalam jangka waktu beberapa saat, serta perubahan jumlah dan pola presipitasi yang terjadi[4]. Presipitasi merupakan endapan air dari atmosfer pada permukaan bumi dalam bentuk cair (tetes hujan) dan padat (es). Bentuk presipitasi yang umum dikenal adalah hujan (rain), gerimis (drizzle), salju (snow), dan batu es (hail). Di wilayah tropis seperti Indonesia, presipitasi lebih didefinisikan sebagai hujan karena sangat jarang dalam bentuk es[5]. Banyaknya dampak negatif yang disebabkan oleh perubahan iklim yang terjadi sehingga diperlukan kegiatan mitigasi bencana untuk mengurangi dampak yang akan sangat besar jika tidak dibuat alur pada tahun-tahun sebelumnya dan adaptasi (dalam bentuk strategi pembangunan yang dapat mereduksi dampak negatif perubahan iklim). Tidak dapat diragukan lagi bahwa kota-kota besar akan sangat besar menyumbang peranan yang cukup penting dalam pembentukan ekonomi yang stabil pada negara-negara berkembang. Kota-kota besar tersebut juga sedang dan akan menghadapi banyak permasalahan yang terkait dengan daya dukung dan kerusakan lingkungan fisik yang terjadi karena kesalahan dari sisi manusia nya atau dari sisi alam. Pertumbuhan populasi yang meningkat tiap tahun dan tidak didukung dengan infrastruktur yang memadai maka kerentanan terhadap dampak perubahan iklim akan semakin tinggi.



Gambar 1. Pembagian Iklim

Fokus besar dalam penelitian kali ini adalah *machine learning* yang dapat menerapkan bagaimana mendeteksi secara berkala pada pola kompleks dan membuat keputusan cerdas berdasarkan data yang tersedia pada tahun-tahun sebelumnya[6]. *Machine learning* dapat mempelajari pola dari data historis yang terdapat pada *database* tahun-tahun sebelumnya untuk memprediksi curah hujan dan banjir selama beberapa bulan bahkan beberapa tahun yang lalu dalam sebuah wilayah dengan beberapa kriteria. Prediksi curah hujan sudah banyak yang melakukan dengan beberapa kriteria parameter, hal tersebut dilakukan oleh beberapa peneliti diantaranya, prediksi curah hujan di kota medan menggunakan metode *backpropagation neural network*[7], prediksi curah hujan bulanan menggunakan jaringan syaraf tiruan dengan beberapa Fungsi Pelatihan *Backpropagation*[8]. Selanjutnya *backpropagation neural network* disebut sebagai BP-NN. Prediksi banjir dengan *machine learning* menggunakan *support vector regression* (SVR) yang pernah dilakukan adalah prediksi tinggi muka air (TMA) Untuk deteksi dini bencana banjir Menggunakan SVR-TVIWPSO[9]. Sebagai suatu rangkaian proses, *machine learning* dapat dibagi menjadi beberapa tahap proses. Tahapan-tahapan tersebut bersifat interaktif, pemakai terlibat langsung atau dengan perantara *knowledge base*. *Machine learning* memiliki makna berupa metode yang digunakan untuk membuat program yang bisa belajar dari data yang tersedia dari beberapa sumber yang dapat dipercaya untuk digunakan data tersebut. Berbeda dengan program komputer biasa dilakukan secara statis karena sudah ada ketentuan apa yang harus dilakukan terlebih dahulu, namun berbeda dengan program *machine learning* merupakan sebuah program yang dirancang untuk mampu belajar sendiri dengan data yang telah tersedia.

Naive Bayes merupakan salah satu algoritma yang dimiliki oleh metode *machine learning* yang dapat digunakan untuk mengklasifikasikan data yang telah tersedia. Penggunaan Naive Bayes dalam penelitian kali ini digunakan untuk mempelajari pola deteksi bencana banjir yang terjadi pada tahun sebelumnya untuk dapat memprediksi seberapa besarnya efek banjir tersebut untuk tahun selanjutnya. Berdasarkan latar belakang ini maka melalui Naive Bayes dapat memprediksi dengan baik data bencana banjir dengan beberapa parameter untuk mengetahui seberapa luas daerah yang masuk dalam kategori Banjir dan kategori Tidak Banjir pada wilayah Kabupaten Bandung. Naive Bayes efektif untuk menentukan apakah bencana banjir ini termasuk

dalam nilai accuracy yang tinggi untuk masuk kategori Banjir atau Tidak Banjir. Parameter yang digunakan dalam penelitian ini adalah kecamatan, intensitas curah hujan, debit air, luas wilayah, kurun waktu hujan, kepadatan penduduk, dan target. Metode Naive Bayes merupakan salah satu metode yang masuk dalam kategori algoritma klasifikasi yang efektif, mudah, efisien dan handal dalam mengatasi data seperti atribut yang kurang atau hilang beserta jumlah data yang digunakan tidak perlu terlalu banyak. Pengklasifikasi bayesian adalah pengklasifikasian yang berupa statistik dan didasarkan pada teorema bayes. Teori keputusan bayes [10] adalah pendekatan data statistik yang paten dalam pengenalan pola (*pattern recognition*), penggunaan algoritma ini berdasarkan dalam hal klasifikasi harus mempunyai masalah yang bisa dilihat statistiknya untuk menghasilkan sebuah nilai. Misalkan X adalah set atribut data dan h kelas variabel dan jika kelas memiliki hubungan dengan atribut maka diperlukan X dan h sebagai variabel acak dan menangkap hubungan peluang $P(h|X)$ ini peluang posterior untuk h dan sebaliknya prior $P(h)$.

Teorema bayes merupakan dasar aturan dari naive bayes classifier berikut teorema bayes akan disajikan pada persamaan (1).

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

Dimana X merupakan data tuple hasil dari pengujian suatu set data yang telah ditentukan masuk ke dalam kelas tertentu yang telah di buat kategorinya. H merupakan suatu hipotesis yang akan menentukan X masuk ke dalam kelas C. $P(H|X)$ merupakan peluang atau probabilitas X yang merupakan data tuple atau bukti yang diperoleh pada saat observasi masuk ke dalam kelas H, dengan kata lain mencari probabilitas X dimiliki oleh kelas H. $P(H|X)$ merupakan probabilitas posterior, H dikondisikan pada X. Sebaliknya $P(H)$ merupakan probabilitas prior, atau probabilitas sebelumnya. Kemudian $P(X|H)$ merupakan probabilitas posterior dimana X dikondisikan pada H. Sedangkan $P(X)$ merupakan probabilitas sebelumnya dari X [11].

II. METODE PENELITIAN

Pada penelitian ini dimulai dengan melakukan pengumpulan informasi dari suatu data yang ingin di angkat pada judul ini dan diambil pada beberapa website resmi, seperti BMKG, BPS, dan BNPD kabupaten bandung. pada penelitian ini berupa data bencana banjir yang ada di daerah kabupaten bandung, Selain itu penggunaan teknik *machine learning* atau pembelajaran mesin dapat digunakan untuk meningkatkan ketangguhan dan kesiapsiagaan data yang didapatkan terhadap bencana banjir,. Penggunaan model pembelajaran mesin untuk memprediksi banjir secara spasial merupakan bidang riset yang disarankan untuk penelitian selanjutnya. Metode *machine learning* telah diterapkan dalam penilaian resiko, dampak banjir, dan dampak iklim yang khususnya dalam pengembangan langkah-langkah mitigasi, persiapan tanggap darurat dan perencanaan pemulihan banjir. Penelusuran dilakukan dari berbagai sumber seperti buku, jurnal, website tertentu, khususnya penelusuran melalui internet, Tujuannya adalah dididapatkan gambaran terkini penelitian dan penerapan upaya mitigasi bencana banjir. Penelitian ini bertujuan untuk mengetahui hasil nilai accuracy yang di tampilkan jika penelitian tersebut menggunakan algoritma Naïve Bayes dalam mendeteksi bencana banjir masuk dalam kategori banjir atau tidak banjir menggunakan teknik data mining. Untuk menjalankan eksperimen kinerja algoritma, digunakan tool *rapid miner*. *Rapid miner* merupakan sebuah perangkat lunak yang bersifat terbuka open source. *Rapid miner* merupakan salah satu solusi untuk melakukan analisis terhadap data mining, text mining dan analisis prediksi dengan berbagai data yang sesuai dengan kebutuhan. *Rapid miner* digunakan dalam berbagai teknik deskriptif dan prediksi untuk menambah wawasan kepada pengguna sehingga dapat membuat keputusan yang paling baik dari beberapa parameter yang digunakan. *Rapid miner* memiliki kurang lebih 500 operator data mining, termasuk operator untuk input, output, data preprocessing dan visualisasi dari model yang di hasilkan dari data yang digunakan. *Rapid miner* merupakan software yang berdiri sendiri untuk analisis data dan sebagai mesin data mining yang dapat diintegrasikan pada jenis data sendiri. *Rapid miner* ditulis dengan menggunakan bahasa java sehingga dapat bekerja di semua sistem operasi. *Rapid miner* sebagai *software open source* yang digunakan untuk metode data mining tidak perlu diragukan lagi karena *software* ini sudah terkemuka di hampir dunia baik di negara maju maupun negara berkembang. *Rapid miner* menempati peringkat pertama sebagai salah satu *software* data mining pada debat terbuka yang dilakukan oleh *KDnuggets*, dan sebuah portal data mining pada tahun kisaran 2010-2011. *Rapid miner* menyediakan *GUI Graphic User Interface* untuk merancang sebuah pipeline analitis. GUI ini akan menghasilkan file *XML Extensible Markup Language* yang mendefinisikan proses analitis keinginan pengguna untuk diterapkan ke data.file ini kemudian dibaca oleh *Rapid miner* untuk menjalankan analis secara otomatis sebelum hal tersebut mesti menampilkan model yang di buat sesuai dengan kebutuhan algoritma yang ingin di gunakan.

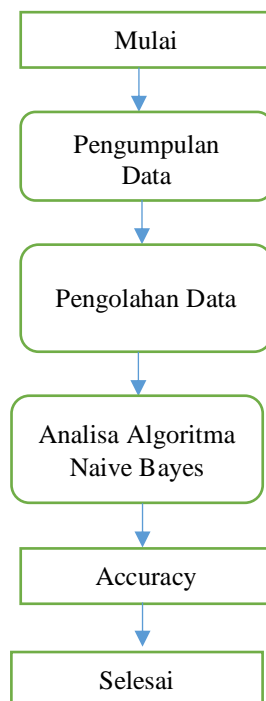
Berdasarkan tugasnya [12] data mining dapat dikelompokkan menjadi beberapa bagian utama yang menjadi tombak dari metode yang ingin di olah, antara lain: deskripsi, estimasi, prediksi, klasifikasi, clustering dan asosiasi. Metode klasifikasi yang digunakan pada penelitian ini adalah algoritma klasifikasi dari data mining yang digunakan adalah algoritma naive bayes, klasifikasi tersebut merupakan bagian dari algoritma data mining yang mempunyai kualifikasi data tertentu untuk dapat digunakan untuk menghasilkan hasil yang dibutuhkan. Klasifikasi ini adalah suatu algoritma yang menggunakan data testing/data training dengan target yang berbentuk (class/label) yang berupa nilai kategorikal/nominal. Contoh dari klasifikasi adalah menganalisa apakah seseorang akan membeli komputer atau tidak dengan beberapa parameter yang digunakan, dalam data mining klasifikasi dapat menghasilkan sebuah prediksi yang dapat digunakan sebagai referensi untuk memutuskan sesuatu hal. Algoritma *naive bayes* mengestimasi peluang-peluang yang terjadi pada kelas bersyarat dengan mengasumsikan bahwa atribut yang ada adalah independen secara bersyarat yang diberikan dengan label kelas label kelas label kelas y.

Tahapan dari penggunaan algoritma naive bayes, sebagai berikut:

1. Menyiapkan data training
2. Setiap data dipresentasikan sebagai vektor berdimensi-n yaitu $X=(x_1,x_2,x_3,\dots,x_n)$
3. n adalah gambaran dari ukuran yang dibuat di test dari n atribut yaitu A_1,A_2,A_3,\dots,A_n
4. M adalah kumpulan kategori yaitu C_1,C_2,C_3,\dots,C_m
5. Diberikan data test X yang tidak diketahui kategorinya, maka classifier akan memprediksi bahwa X adalah milik kategori dengan posterior probability tertinggi berdasarkan kondisi X
6. Naive bayes classifier menandai bahwa test X yang tidak diketahui tadi ke kategori C1 jika

Confusion matrix merupakan sebuah tabel yang digunakan untuk menggambarkan kinerja model dari klasifikasi yang digunakan (atau "*classifier*") pada set data uji yang nilai sebenarnya diketahui namun di keluarkan dalam bentuk tabel. Hal ini memungkinkan identifikasi yang mudah dari kesukaran antar *class*, misal satu kelas umumnya salah diberi label sebagai yang lain benar di berikan labelnya. Jumlah prediksi yang benar dan salah dirangkum dengan nilai-nilai hitung dan dipecah oleh masing-masing *class*. *Confusion matrix* menunjukkan cara-cara ketika klasifikasi bingung menentukan *class*-nya dalam membuat prediksi. Ini memberi informasi detail tidak hanya tentang kesalahan yang dibuat oleh classifier tetapi lebih penting lagi jenis kesalahan yang terjadi dari data-data yang dihasilkan.

Tahapan dari proses penelitian ini dituangkan pada gambar 2.



Gambar 2. Tahapan Penelitian

A. Pengumpulan Data

Pada penelitian ini data yang digunakan merupakan data sekunder yaitu data yang diambil dari website resmi BMKG Indonesia, dengan beberapa data pendukung yang diambil dari website resmi Badan Nasional Penanggulangan Daerah (BNPD), dan Badan Pusat Statistik (BPS) Indonesia. Data set yang diperoleh kurang lebih sejumlah 2000 data dalam kurun waktu tahun 2010 sampai dengan 2013, namun untuk data testing yang akan di implementasikan ke tools *rapid minner* hanya 800 data. Contoh data set yang digunakan pada gambar 3, di bawah ini:

Kecamatan	Intensitas Curah Hujan	Debit Air	Luas Wilayah	Kurun Waktu Hujan	Kepadatan Penduduk	target
Banjaran	115	220	48.47	1	71309	Tidak Banjir
Banjaran	196	345	48.47	1	71309	Tidak Banjir
Banjaran	175	210	48.47	1	71309	Tidak Banjir
Banjaran	108	306	48.47	4	71309	Tidak Banjir
Banjaran	174	203	48.47	3	71309	Tidak Banjir
Banjaran	70	156	48.47	1	71309	Tidak Banjir
Banjaran	77	104	48.47	3	71309	Tidak Banjir
Banjaran	108	171	48.47	3	71309	Tidak Banjir
Banjaran	451	212	48.47	5	71309	Banjir
Bojongsoang	123	2.77	27.81	2	98239	Tidak Banjir
Bojongsoang	198	3.97	27.81	1	98239	Banjir
Bojongsoang	186	4.01	27.81	4	98239	Tidak Banjir
Bojongsoang	108	4.91	27.81	4	98239	Tidak Banjir
Bojongsoang	176	3.56	27.81	3	98239	Tidak Banjir
Bojongsoang	60	3.2	27.81	4	98239	Tidak Banjir

Gambar 3. Contoh Data Set

B. Pengolahan Data

Tahap ini data yang telah di kumpulkan dari website yang dituju untuk mendapatkan beberapa data yang di butuhkan masuk ke dalam tahap pembersihan data yang tidak terpakai, data yang ada non value, dan ketidak konsistenan dalam penulisan data.

C. Analisa Algoritma Naive Bayes

Analisa data dari data yang sudah di olah menjadi data yang siap di gunakan untuk di masukkan ke dalam tools Rapidminner dengan model-model yang sesuai dengan ketentuan dari algoritma *Naive Bayes*

D. Accuracy

Hasil yang di dapatkan dari model yang di buat menggunakan algoritma Naive Bayes dengan memasukkan data yang sudah di filter pada microsoft excel.

Khusus untuk data training ada hal yang harus di lakukan yaitu dengan memasukan beberapa sampel klasifikasi Naive Bayes yang digunakan adalah metode cross validation dengan parameter 10 folds. Sementara itu untuk parameter Naive Bayes digunakan metode polinomial. Pemilihan metode tersebut didasarkan dari hasil yang ingin di cari yaitu data bencana banjir yang bersifat binomial yaitu berupa keluaran yang bersifat kategorikal dalam kasus ini output yang spesifik hanya bernilai dua kondisi yang dihasilkan yaitu banjir dan tidak banjir. Untuk mendekati hal tersebut maka digunakan metode polinomial pada klasifikasi Naive Bayes. Setelah data training digunakan maka selanjutnya parameter dari hasil klasifikasi berupa Kecamatan, Intensitas Curah Hujan, Debit Air, Luas Wilayah, Kurun Waktu Hujan, Kepadatan Penduduk, dan target yang akan digenerate dalam bentuk HTML berupa report untuk menghasilkan nilai akurasi, dan recall.

III.HASIL DAN PEMBAHASAN

Hasil dari penelitian yang di lakukan ini yaitu menggunakan algoritma Naive Bayes dengan tools *Rapidminner*, yang menggunakan data set kurang lebih 2000 data namun untuk data training yang di uji coba pada tools Rapidminner untuk mendapatkan hasil accuracy yang di butuhkan sesuai dengan algoritma yang ingin di gunakan kurang lebih 800 data, dengan pembagian data training 500 *record* data dan 300 *record* data untuk data testing untuk dapat menghasilkan sebuah model dari *rapidminner*. Hasil nilai *accuracy* dari algoritma *naive*

bayes yang di gunakan menggunakan tools *rapidminer*, menghasilkan 76.73% sesuai dengan gambar di bawah ini:

accuracy: 76.73% +/- 23.13% (micro average: 76.19%)

	true Tidak Banjir	true Banjir	class precision
pred. Tidak Banjir	76	8	90.48%
pred. Banjir	22	20	47.62%
class recall	77.55%	71.43%	

Gambar 4. Accuracy Naive Bayes

Gambar 4 merupakan perhitungan accuracy data *testing* menggunakan algoritma Naive Bayes yang menghasilkan nilai accuracy nya 76.73%. Diketahui hasil yang di hasilkan tersebut di datapat dari data *testing* terdiri dari 126 *record* data yang sudah di filter dari data set asli yang di tarik dari website resmi sumber yang terkait.

PerformanceVector

```

PerformanceVector:
accuracy: 76.73% +/- 23.13% (micro average: 76.19%)
ConfusionMatrix:
True:   Tidak Banjir   Banjir
Tidak Banjir:   76       8
Banjir:  22          20
precision: 63.66% +/- 34.48% (micro average: 47.62%) (positive class: Banjir)
ConfusionMatrix:
True:   Tidak Banjir   Banjir
Tidak Banjir:   76       8
Banjir:  22          20
recall: 73.83% +/- 26.08% (micro average: 71.43%) (positive class: Banjir)
ConfusionMatrix:
True:   Tidak Banjir   Banjir
Tidak Banjir:   76       8
Banjir:  22          20
AUC (optimistic): 0.801 +/- 0.189 (micro average: 0.801) (positive class: Banjir)
AUC: 0.801 +/- 0.189 (micro average: 0.801) (positive class: Banjir)
AUC (pessimistic): 0.801 +/- 0.189 (micro average: 0.801) (positive class: Banjir)
    
```

Gambar 5. Performance Vector

Gambar 5 merupakan perhitungan accuracy data menggunakan algoritma Naive Bayes. Diketahui data testing terdiri dari 126 *record* data, 20 data diprediksikan BANJIR kenyataannya benar BANJIR, 8 data diprediksikan BANJIR dalam kenyataannya TIDAK BANJIR, 22 data diprediksikan TIDAK BANJIR namun dalam kenyataannya BANJIR dan 76 data diprediksikan TIDAK BANJIR pada kenyataan benar-benar TIDAK BANJIR.

IV. KESIMPULAN

Penelitian ini berhasil mengimplementasikan metode klasifikasi dengan penggunaan algoritma *naive bayes* untuk penerapan deteksi bencana banjir yang terjadi pada beberapa tahun sebelumnya untuk dapat memitigasikan bencana banjir di tahun-tahun selanjutnya dengan menggunakan tools Rapidminer, dan didapatkan tingkat kecocokan / akurasi sebesar 76.73%.

DAFTAR PUSTAKA

- [1] Amril Mutoi Siregar. PENERAPAN ALGORITMA K-MEANS UNTUK PENGELOMPOKAN DAERAH RAWAN BENCANA DI INDONESIA. Universitas Buana Perjuangan Karawang
- [2] Rahayu, Harkunti P., I. I. Wahdiny, A. Utami dan M. Asparini. 2009. Banjir dan Upaya Penanggulangannya. Bandung : PROMISE Indonesia
- [3] Ligal, S. 2008. Pendekatan Pencegahan dan Penanggulangan Banjir. Jurnal. Dinamika Teknik Sipil Volume 8, No. 2 Juli 2008
- [4] FoxNews,(<http://www.foxnews.com/story/2007/09/04/nasa-global-warming-tocausemore-severe-tornadoes-storms.html>), diakses tanggal 10 Desember 2017.
- [5] Bayong, T. H. K. 1999. Klimatologi Umum. Bandung: ITB.
- [6] Purnamasari, Detty, Jonathan H, Yoga Perdana S, Fuji Ihsani, and I wayan S Wicaksana. Get Easy Using Weka. Jakarta Timur: Dapur Buku, 2013.
- [7] Yudhi Andrian, Erlinda Ningsih. "Prediksi Curah Hujan Di Kota Medan Menggunakan Metode Backpropagation Neural Network." Seminar Nasional Informatika. Medan: e-joynal potensi utama, 2014.
- [8] Cici Oktaviani, Afdal. "Prediksi Curah Hujan Bulanan Menggunakan Jaringan Syaraf Tiruandengan beberapa Fungsi Pelatihan Backpropagation." Jurnal fisika Unand, 2014: 228-237.
- [9] Arief Andy Soebroto, Imam Cholissodin, Randy Cahya Wihandika, Ziya El Arief. "Prediksi Tinggi Muka Air (TMA) Untuk Deteksi Dini Bencana Banjir Menggunakan SVRTVIWPSO." Jurnal Teknologi Informasi dan Ilmu Komputer, 2015: 79-86
- [10] Santosa, Budi. Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis. Yogyakarta : Graha Ilmu, 2007.
- [11] Han, J., and M.Kamber. 2006. Data Mining Concept and Techniques Second Edition. San Frasisco : Elsevier.
- [12] Larose, Daniel T. Data Mining Methods and Models. Canada : A Jhon Willey & Sons, Inc.,Publication, 2005.