Nama : Dea Tri Ananda

NIM : 09031181621131

Jurusan : Sistem Informasi

**ANALISIS PEMANFAATAN**

**KNOWLEDGE DISCOVERY IN DATABASES (KDD)**

**UNTUK MENGETAHUI PRAKIRAAN CUACA**

1. **PENDAHULUAN**

*Knowledge Discovery in Databases* (KDD) merupakan keseluruhan proses non-trivial untuk mencari dan mengidentifikasi pola (pattern) dalam data, dimana pola yang ditemukan bersifat sah, baru, dapat bermanfaat dan dapat dimengerti. *Knowledge Discovery in Databases* adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam data berukuran besar. Adapun beberapa tahapan proses *Knowledge Discovery in Databases* yakni sebagai berikut:

1. *Data Selection*

Pemilihan data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam KDD dimulai.

1. *Data Cleaning*

Sebelum proses data mining dilakukan, maka perlu dilakukan proses cleaning pada data yang menjadi fokus KDD. Proses ini mencakup membuang duplikasi data, memeriksa data yang inkosiste, dan memperbaiki kesalahan pada data.

1. *Data Integration*

Data yang diperlukan untuk data mining tidak hanya berasal dari satu database tetapi juga berasal dari beberapa database ataupun file teks, hasil integrasi data biasanya dihasilkan dalam bentuk data warehouse.

1. *Data Transformation*

Merupakan proses pencarian fitur-fitur yang berguna untuk mempresentasikan data bergantung pada goal yang ingin dicapai. Data yang telah dipilih pada proses transformasi merupakan data yang telah sesuai untuk proses data mining.

1. *Data Mining*

Tahap ini merupakan proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma dalam mining sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan.

1. *Interpretation/Evaluation*

Tahap ini merupakan bagian dari proses KDD yang mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesa sebelumnya.

1. *Knowledge Presentation*

Tahap ini merupakan tahap penggunaan teknik-teknik visualisasi dan representasi untuk menyajikan pengetahuan yang telah diperoleh kepada user.

1. **DATASET PREDIKSI CUACA DARI SITUS KDD**

Dataset merupakan sekumpulan data yang saling berkaitan satu sama lain yang bersifat spesifik terhadap suatu kasus, missalnya dalam tulisan ini berupa dataset prediksi cuaca di tahun 2015 di sebuah negara. Dataset dapat direprentasikan dalam berbagai bentuk seperti tabel dalam basis data, bentuk matriks, teks, *Comma Separated Value* (CSV), dan sebagainya. *Dataset* dapat dikumpulkan dan dibentuk oleh seseorang, sekelompok bahkan suatu organisasi dan dipublikasikan secara *online* seperti misalnya situs Kaggle (www.kaggle.com) yang berisi berbagai macam *dataset*.

Pada tulisan ini, penulis mendapatkan informasi dataset dari situs Kaggle “[www.kaggle.com](http://www.kaggle.com)” tentang Data Cuaca di tahun 2015. Dimana dataset ini menjelaskan menjelaskan berbagai informasi cuaca di tahun 2015. Format dataset telah divirtualisasi ke dalam format excel. Dataset tersebut terdiri dari 368 baris, dan 27 kolom. Kolomnya terdiri dari Date/time, Year, Month, Day, Data Quality, Max Temp (°C), Max Temp Flag, Min Temp (°C), Min Temp Flag, Mean Temp (°C), Mean Temp Flag, Heat Deg Days (°C), Heat Deg Days Flag, Cool Deg Days, Cool Deg Days (°C), Cool Deg Days Flag, Total Rain (mm), Total Rain Flag, Tolat Snw (cm), Total Snow Flag, Total Percip (mm), Total Percip Flag, Snow on Grnd (cm), Snow on Grnd Flag, Dir of Max Gust (10s deg), Dir of Max Gust Flag, Spd of Max gust (Km/h), Spd of Max Gust Flag.

Dataset ini dibuat untuk mengklasifikasikan hari dengan kelembaban rendah vs hari tidak lembab (yaitu, hari dengan kelembaban normal atau tinggi), variabel yang termasuk adalah pengukuran cuaca di pagi hari, dengan satu pengukuran, yaitu relatif lembab, di sore hari. Idenya adalah menggunakan nilai cuaca pagi untuk memprediksi apakah hari akan kelembaban rendah atau tidak berdasarkan pengukuran sore kelembaban relatif.

Dari data-data yang ada tersimpan berbagai informasi yang dapat dipergunakan untuk memprediksi cuaca dimasa mendatang. Untuk itu diperlukan Knowledge Discovery in Databases (KDD) untuk mengetahui pola dalam menghasilkan informasi yang dapat berguna dimasa mendatang.

1. **ANALISIS JURNAL PENDUKUNG INFORMASI**
2. **Jurnal “PENERAPAN DATA MINING UNTUK PRAKIRAAN CUACA DI KOTA MALANG MENGGUNAKAN ALGORITMA *ITERATIVE DICHOTOMISER* TREE (ID3)”.**

Analisis:

Dalam persiapan informasi ramalan cuaca ada beberapa kendala yang meilbatkan banyak orang, sumber data dan ramalan cuaca bergantung pada kemampuan pelopor, sehingga interpretasi yang dihasilkan mungkin berbeda antara peramal karena pengalaman mereka sendiri. Perbedaan interpretasi dapat membingungkan pengguna dan berpotensi mengurangi kualitas informasi diajukan. Dalam jurnal ini digunakan model ramalan dengan data penambangan menggunakan algoritma ID3 untuk mendapatkan model yang sesuai sehingga memudahkan proses analisis dan ramalan cuaca. Data-data yang dianalisis dari jurnal ini diperoleh dari situs web BMKG dan data yang diambil adalah data cuaca pada tanggal 1 Januari 2012 untuk total 2414 data dan 9 atribut. Dalam proses pengujian dilakukan dengan 179 data (10% dari kumpulan data) acak, diketahui bahwa ada 112 database sesuai dengan data cuaca yang sebenarnya. Dari jurnal ini dapat diketahui kesimpulan bahwa tingkat akurasi sebesar 73,74%. Atribut Kelembaban memiliki pengaruh paling dominan dalam menentukan prediksi cuaca. Kelas tidak hujan cenderung muncul lebih banyak daripada kelas lain seperti Hujan Ringan, Hujan Sedang, Hujan Lebat, dan Hujan Sangat Lebat. (Tp and Sari 2017)

1. **Jurnal “PEMANFAATAN DATA MINING UNTUK PRAKIAAN CUACA”**

Analisis:

Dari jurnal ini, informasi yang diperoleh yaitu Proses prakiraan cuaca memerlukan banyak komponen data cuaca, jumlah data yang besar serta kemampuan prakirawan. Hal ini menyebabkan ketepatan dan kecepatan prakiraan kurang terpenuhi. Untuk memecahkan masalah tersebut, telah dilakukan penelitian model prediksi menggunakan beberapa teknik data mining yakni Association Rule, C4.5, Classification dan Random Forest. Data masukan adalah data sinoptik 9 stasiun maritim tahun 2009. Data masukan tersebut terdiri dari kecepatan angin, tutupan awan, suhu udara dan suhu titik embun. Data untuk pengujian model adalah data sinoptik Stasiun Meteorologi Maritim Tanjung Priok sejak tahun 2002 hingga 2010. Dari serangkaian pembuatan, pemilihan dan pengujian model, hasil penelitian menunjukkan Association Rule mempunyai tingkat akurasi 60.9%, penulis memilih *rule-rule* dengan bentuk asosiasi jika "komponen cuaca seperti suhu udara atau suhu titik embun atau tutupan awan atau kecepatan angin mempunyai nilai tertentu" maka "kondisi hujan yang terjadi bagaimana". Dengan hasil penyaringan didapatkan satu model dengan bentuk : **Jika suhu udara > 23.9 dan suhu titik embun 23.6 maka terjadi hujan**, sedangkan C4.5 mempunyai tingkat akurasi 68.5% dengan faktor penentu hujan atau tidak hujan adalah suhu udara, suhu titik embun, dan tutupan awan. Namun interval syarat batasnya tidak jauh berbeda. Maka dapat ditarik kesimpulan komponen cuaca yang dominan memungkinkan terjadinya hujan adalah suhu udara, suhu titik embun, dan tutupan awan.

1. **Jurnal “PERANCANGAN SISTEM PERAMALAN CUACA BERBASIS LOGIKA FUZZY”**

Analisis:

Dari jurnal ini, informasi yang diperoleh yakni dalam menentukan sebuah peramalan khususnya cuaca yang penting ditentukan adalah fungsi keanggotaan dan rule yang digunakan. Logika Fuzzy dengan menggunakan metode sugeno sangat baik di gunakan dalam peramalan karena tingkat keakuratan diatas 60%. Semakin banyak variabel-variabel yang dijadikan input maka akan menghasilkan output yang semakin baik (akurat). Variabel input yang mempengaruhi kondisi cuaca yaitu temperatur atau suhu, kelembaban, dan kecepatan angin. Adapun rencana rule base prediksi cuaca khususnya hujan yang penulis lakukan seperti pada tabel dibawah ini:





Dari tabel diatas dapat diperoleh informasi berupa jika suhu dingin, kelembaban lembab dan angin pelan maka terjadi hujan sedang. Dan jika suhu dingin, kelembaban lembab dan angin sedang maka terjadi hujan sedang, Dan jika suhu dingin, kelembaban lembab dan angin kencang maka terjadi hujan deras. Dan jika suhu dingin, kelembaban sedang dan angin pelan maka terjadi hujan sedang. Dan jika suhu dingin, kelembaban sedang dan angin sedang maka terjadi hujan sedang. Dan jika suhu dingin, kelembaban lembab dan angin kencang maka terjadi hujan deras. Dan jika suhu dingin, kelembaban tidak lembab dan angin pelan maka terjadi hujan rendah. Dan jika suhu dingin, kelembaban tidak lembab dan angin sedang maka terjadi hujan rendah. Dan jika suhu dingin, kelembaban tidak lembab dan angin kencang maka terjadi hujan rendah.

1. **Jurnal “PENERAPAN LOGIKA FUZZY UNTUK MEMPREDIKSI CUACA HARIAN DI BANJARBARU”**

Analisis:

Dari jurnal tersebut, informasi yang diperoleh yaitu prediksi cuaca hujan ringan dan hujan lebat dengan model logika *fuzzy* didapatkan hasil yang layak yakni dibawah 65%. Hal tersebut dikarenakan pada proses pertumbuhan awan tidak selalu diakhiri dengan terjadinya hujan. Jatuhnya butiran air sebagai hujan dipengaruhi beberapa faktor yaitu jenis awan, tinggi dasar awan, jumlah uap air dan inti kondensasi. Semakin banyak kandungan uap air di uadara, peluang terjadinya hujan akan semakin besar.

1. **Jurnal “PREDIKSI HUJAN UNTUK WILAYAH KOTA KUPANG DENGAN METODE *DATA MINING* POHON KEPUTUSAN C4.5”**

Analisis:

Dari jurnal tersebut, informasi yang diperoleh yaitu informasi cuaca sudah menjadi suatu kebutuhan masyarakat saat ini, sehingga diperlukan prakiraan yang tepat, akurat dan mudah dipahami. Dalam melakukan prakiraan cuaca banyak sekali metode dan data yang digunakan. Salah satunya data yang dapat digunakan adalah data stabilitas udara yang bisa didapat dari pengamatan pilot balon dan radiosonde.

Dalam jurnal ini data yang digunakan adalah data radiosonde sebagai tambahan referensi dalam membuat prakiraan khususnya potensi kejadian hujan. Penulis mencari nilai ambang batas baru dari indeks stabilitas yang lebih sesuai untuk kota kupang sebagai daerah penelitian dalam memprediksi potensi kejadian hujan. Dan indeks yang dicari juga dibagi menjadi masing-masing indeks dan indeks gabungan. Kemudian nilai ambang batas baru yang didapat diverifikasi dengan menggunakan metode *dichotomous.* Dan didapat nilai ambang batas baru dengan nilai verifikasi yang baik pada bulan Februari jam 00.00 UTC untuk nilai TPW, jam 12.00 UTC untuk indeks SI. Pada bulan Desember jam 00.00 UTC nilai TPW, KI, SWEAT, SI, sedangkan jam 12.00 UTC untuk nilai SWEAT. Dan untuk data tahunan didapat nilai ambang batas baru pada jam 00.00 UTC untuk nilai TPW dan KI. Dan tingkat akurasi dari nilai ambang batas tersebut paling rendah 74% dan yang paling baik 88%.

Pada KI formulasi yang digunakan adalah perhitungan kondisi suhu parsel udara, suhu udara lingkungkan, dengan kondisi suhu titik embun. Suhu titik embun ini menggambarkan kondisi kelembapan udara.

Pada SWEAT formulasi yang digunakan adalah perhitungan kondisi suhu parsel udara, suhu udara lingkungan, kondisi titik embun dan ditambah kondisi arah dan kecepatan angin. Selain adanya perhitungan suhu titik embun yang menggambarkan kondisi kelembapan suatu udara, indeks ini juga memperhitungna pola pergerakan angin yang dapat menggambarkan pola pembentukan awan atau perluruhan awan akibat gerakan udara.

Pada TPW formulasi yang digunakan adalah menghitung jumlah uap air yang tersedia pada kolom udara suatu lokasi pengamatan.

Dari ketiga indeks tersebut semuanya memuat perhitungan kondisi uap air sebagai perhitungan utama dalam pembentukan hujan. Inilah kenapa ketiga indeks ini banyak didapat dalam mencari nilai ambang batas indeks untuk kejadian hujan.

1. **KESIMPULAN**

Berdasarkan analisis yang tekah dilakukan di ke-5 jurnal diatas maka dapat ditarik kesimpulan bahwa untuk mengetahui pola cuaca terutama terjadinya hujan dapat diketahui menggunakan berbagai metode. Dan juga dari metode-mtode tersebut dapat ditemukan beberapa pola untuk memprediksi cuaca yang akan terjadi yaitu jika "komponen cuaca seperti suhu udara atau suhu titik embun atau tutupan awan atau kecepatan angin mempunyai nilai tertentu" maka "kondisi hujan yang terjadi bagaimana". Jika yang digunakan adalah Association Rule maka pola yang diperoleh yakni **Jika suhu udara > 23.9 dan suhu titik embun 23.6 maka terjadi hujan**.

Selain itu juga didapat pola seperti jika suhu dingin, kelembaban lembab dan angin pelan maka terjadi hujan sedang. Dan jika suhu dingin, kelembaban lembab dan angin sedang maka terjadi hujan sedang, Dan jika suhu dingin, kelembaban lembab dan angin kencang maka terjadi hujan deras. Dan jika suhu dingin, kelembaban sedang dan angin pelan maka terjadi hujan sedang. Dan jika suhu dingin, kelembaban sedang dan angin sedang maka terjadi hujan sedang. Dan jika suhu dingin, kelembaban lembab dan angin kencang maka terjadi hujan deras. Dan jika suhu dingin, kelembaban tidak lembab dan angin pelan maka terjadi hujan rendah. Dan jika suhu dingin, kelembaban tidak lembab dan angin sedang maka terjadi hujan rendah. Dan jika suhu dingin, kelembaban tidak lembab dan angin kencang maka terjadi hujan rendah.

**DAFTAR PUSTAKA**

TP, B. P., & Sari, R. D. I. (2017). PENERAPAN DATA MINING UNTUK PRAKIRAAN CUACA DI KOTA MALANG MENGGUNAKAN ALGORITMA ITERATIVE DICHOTOMISER TREE (ID3). *Joutica*, *2*(2).

Mujiasih, S. (2011). Pemanfatan Data Mining Untuk Prakiraan Cuaca. *Jurnal Meteorologi dan Geofisika*, *12*(2).

Puspita, E. S., & Yulianti, L. (2016). Perancangan sistem peramalan cuaca berbasis logika fuzzy. *Jurnal Media Infotama*, *12*(1).

Mahanani, U., Fahrudin, A. E., & Nurlina, N. (2016). Penerapan Logika Fuzzy Untuk Memprediksi Cuaca Harian di Banjarbaru. *Jurnal Fisika Flux*, *12*(1), 13-19.

W.Wardhana, I Ketut. (2016). PREDIKSI HUJAN UNTUK WILAYAH KOTA KUPANG DENGAN METODE *DATA MINING* POHON KEPUTUSAN C4.5, 8.