LAPORAN TUGAS AKHIR DATA MINING

IMPLEMENTASI MODEL CRISP-DM MENGGUNAKAN METODE K-NEAREST NEIGHBOR UNTUK MEMPREDIKSI PENYEBARAN DEMAM BERDARAH DI KOTA BANDUNG

Laporan ini Disusun Sebagai Tugas Akhir Mata Kuliah Data Mining

DOSEN PENGAMPU:

ABU SALAM, M.Kom

DISUSUN OLEH:

Dhany Septiandhika Pratama

(A11.2019.11750)

UNIVERSITAS DIAN NUSWANTORO SEMARANG

FAKULTAS ILMU KOMPUTER

TEKNIK INFORMATIKA

2021/2022

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

BAB I
SUMBER JURNAL

Dalam eksperimen kali ini saya menggunakan 2 paper jurnal penelitian sebagai acuan tugas akhir yaitu yang pertama dengan judul Implementasi CRISP-DM Model Menggunakan Metode Decision Tree dengan Algoritma CART untuk Prediksi Curah Hujan Berpotensi Banjir yang dibuat oleh Msy Aulia Hasanah, Sopian Soim, Ade Silvia Handayani dan paper jurnal yang kedua berjudul Prediksi Penyebaran Kasus DBD Dengan Metode KNN Di Kabupaten Karawang yang dibuat oleh Bimo Aditya Wahyudi, Hannie, Aries Soeharso. Penjelasan jurnal sebagai berikut:

1. Jurnal berjudul **Implementasi CRISP-DM Model Menggunakan Metode Decision Tree dengan Algoritma CART untuk Prediksi Curah Hujan Berpotensi Banjir** dilatarbelakangi oleh banjir yang menjadi salah satu bencana bagi sebagian masyarakat, terutama yang tinggal didaerah dataran rendah maupun dibantaran sungai.(Hasanah et al., 2021) Fenomena banjir ini sangatlah berpengaruh, baik itu dari segi ekonomi, lingkungan maupun keselamatan masyarakat. Faktor penyebab banjir yang paling utama ialah curah hujan. Tingginya intensitas curah hujan dapat mempengaruhi jumlah volume debit air yang mengalir pada saluran sungai yang melebihi kapasitas alirannya. Sehingga aliran sungai tersebut meluap dan menggenangi dataran yang rendah disekitaran dataran banjir. Banjir diakibatkan oleh ketidaksetaraan antara aliran masuk intensitas hujan yang lebih tinggi dari pada aliran keluar, terutama bila drainase saluran air dan daerah resapan tidak berjalan dengan lancar. Menggingat curah hujan salah satu faktor dinamis sebagai penyebab utama banjir, dibutukannya teknologi dan informasi untuk mengelola data. Data tersebut akan kita olah menjadi pengetahuan sebagai acuan dalam membaca serta mengetahui pola pendekatan tersembunyi dari kumpulan data, melakukan analisis tentang pengelompokan antara data dan atribut untuk mendukung pengambilan keputusan serta pembuatan kebijakan dalam memberikan informasi curah hujan yang berpengaruh terhadap segala macam aktifitas seperti keselamatan masyarakat dan sosial-ekonomi. Data ini diolah menjadi sebuah pengetahuan agar dapat bermanfaat bagi banyak orang, dengan mengubah menjadi pengetahuan, manusia

dapat melakukan prediksi dan estimasi tentang apa yang akan terjadi kedepan. Maka dari itu, perlu adanya proses yang menggunakan teknik statistik, matematik, kecerdasan buatan (Artificial Intelegent) dan Machine Learning untuk mengestrak pengetahuan atau menemukan pola dari suatu data yang besar.

Tahapan Penelitian menggunakan metodologi CRISP-DM dengan tahapan:

1) Bussiness Understanding

   Berhubungan langsung dengan data Curah hujan untuk menggali pengetahuan tentang suatu pola terhadap intensitas curah hujan yang berpotensi terhadap bencana banjir. Serta untuk melihat parameter-parameter apa saja yang mempengaruhinya terhadap tingginya intensitas curah hujan tersebut.

2) Data Understanding

   Penelitian tersebut menggunakan data yang berasal dari BMKG Stasiun Klimatologi Kelas l Palembang. Data tersebut merupakan data iklim dari periode 2011-2020. Parameter yang digunakan seperti suhu (T), kelembaban (RH), lama penyinaran matahari (ss) serta curah hujan. Verifikasi terhadap klasifikasi curah hujan yang berpotensi banjir dikategorikan hujan lebat dan hujan sangat lebat berdasarkan nilai ambang (threshold) BMKG.

3) Data Preparation

   Pada Tahap Data preparaton kumpulan data yang akan diubah ke bentuk data yang sesuai dalam proses data mining. Data tersebut dilakukan perubahan dari format xls menjadi format csv, Dataset intensitas hujan lebat diberi label angka 0 (nol) sedangkan intensitas hujan sangat lebat diberi label angka 1 (satu). Setelah itu data tersebut disimpan, dan diimport ke tools data mining menggunakan python.

4) Modelling

   Visualisasi pengolahan data pada algoritma CART menghasilkan 17 nodes dan 18 leaf dengan tingkat kedalaman 13 level. Pada root node memiliki splitting kriteria pada atribut kelembaban rata-rata $\leq$ 88% dengan kata lain disini memanfaatkan features kelembaban sebagai pemecah kondisi sebelum proses splitting. Berikutnya terdapat beberapa informasi seperti perolehan parameter gini impurity yaitu sebesar 0.16 dengan sampel (jumlah data) 112 data untuk kelas 0 (hujan lebat) dan 11 data untuk kelas 1 (hujan sangat lebat).

Setelah dikenakan proses splitting, dari kriteria yang ada pada root node menghasilkan 2 ruas yaitu 1 internal node dan 1 leaf node. Pada leaf node (ruas kanan) menghasilkan sebuah keputusan bila kelembaban rata-rata > 88% maka intensitas curah hujan berpeluang dengan kriteria hujan lebat.

5) Evaluation

Setelah pola klasifikasi didapatkan pada algoritma CART selanjutnya dilakukan tahap evaluasi komparasi algoritma dengan parameter yang digunakan ialah Confusion Matrix yang pada dasarnya untuk memberikan informasi perbandingan hasil yang telah dilakukan oleh model dengan hasil klasifkasi sebenarnya dengan melihat nilai akurasi, presisi dan recall. hasil nilai Accuracy = 0.89, Precision = 0.37 dan recall = 0.27. Sehingga didapatkan tingkat akurasi algoritma CART sebesar 89,4% dengan jumlah prediksi benar 110 data dari total data testing 123 dataset.

6) Deployment

Setelah tahap evaluasi dimana menilai secara detail hasil dari sebuah model maka dilakukan pengimplementasian dari keseluruhan model yang telah dibangun. Selain itu juga dilakukan penyesuaian terhadap model sehingga dapat menghasilkan suatu hasil yang sesuai dengan target awal tahap CRISP-DM ini.

Kesimpulan dari penelitian tersebut, telah dilakukan klasifikasi data curah hujan menggunakan metode decision tree algoritma CART (Classification And Regression Tree) dengan teknik data mining CRISP-DM. Dataset terdiri 6 atribut yaitu suhu ratarata, suhu min, suhu max, kelembaban, lama penyinaran matahari dan curah hujan . Dari 3.653 dataset curah hujan hanya 123 record yang dikotomi curah hujan yang berintensitas hujan lebat dan hujan sangat lebat namun dari hasi pengujian algoritma ini memiliki kinerja yang cukup baik dengan akurasi sebesar 89,4% dengan Evaluasi dan validasi menggunakan parameter uji Confusion Matrix, dari perolehan akurasi tersebut didapatkan bahwa jumlah prediksi benar adalah 110 data dari jumlah total data uji yaitu 123 data.

2. Jurnal berjudul **Prediksi Penyebaran Kasus DBD dengan Metode K-NN di Kabupaten Karawang** dilatarbelakangi oleh Demam Berdarah Dengue (DBD)

merupakan salah satu penyakit epidemik.(Budiman et al., 2021) Kabupaten Karawang merupakan salah satu daerah endemis Demam Berdarah Dengue (DBD) di Jawa Barat, Indonesia. Munculnya penyakit ini berkaitan dengan kondisi wilayah seperti lingkungan dan juga cuaca serta perilaku masyarakat juga berkaitan terhadap penyebaran kasus DBD. Penyakit epidemik seperti DBD merupakan salah satu penyakit berbahaya, sehingga perlu dilakukan prediksi untuk memprediksi wilayah mana saja yang rawan terkena kasus DBD. Melakukan prediksi tersebut bertujuan agar dapat dengan segera mengetahui wilayah mana yang tingkat persebaran paling tinggi hingga paling rendah untuk melakukan tindakan dan pencegahan yang sesuai. Untuk membantu melakukan prediksi tersebut, digunakan metode K-Nearest Neighbor (K-NN) dan K-Fold Cross Validation.

Tahapan Penelitian menggunakan metodologi CRISP-DM dengan tahapan:

1) Bussiness Understanding

   Berdasarkan permasalahan yaitu dengan tingginya kasus demam berdarah yang terjadi di beberapa wilayah pada Kabupaten Karawang. Maka dilakukan tujuan untuk memprediksi daerah yang rawan terjangkit demam berdarah di Kabupaten Karawang dengan mengklasifikasikan data menggunakan teknik pemodelan. Sehingga nantinya dapat membantu instansi terkait sebagai referensi dalam mengantisipasi daerah yang rawan demam berdarah di Kabupaten Karawang dengan meninjau hasil dari data yang telah diolah sebelumnya.

2) Data Understanding

   Pada tahap pengumpulan data, data yang digunakan pada penelitian tersebut bersifat sekunder, dimana data diperoleh melalui situs web Badan Pusat Statistik (BPS) Kabupaten Karawang (https://karawangkab.bps.go.id/) dan Dinas Kesehatan Kabupaten Karawang (https://dinkes.karawangkab.go.id/). Terdapat beberapa data yang telah dikumpulkan pada penelitian tersebut yaitu kecamatan, rata-rata curah hujan, kepadatan penduduk, kelembapan, mobilitas penduduk dan data kasus demam berdarah. Data yang dikumpulkan pada tahun 2016-2020, dimana terdiri dari 30 kecamatan di Kabupaten Karawang

3) Data Preparation

Pada tahap data preparation tersebut mencakup semua kegiatan dalam menyiapkan dan membangun dataset yang akan diterapkan ke dalam tools pemodelan untuk selanjutnya dilakukan proses Data Mining. Dataset yang digunakan pada penelitian ini sebanyak 150 record data dengan menggunakan 5 parameter sebagai input.

4) Modelling

Pengolahan data dilakukan dengan menggunakan tools Rapidminer yang dimana Algoritma KNN dihitung secara otomatis menggunakan tools tersebut. Pada tahap ini dilakukan 5 kali pengujian dengan nilai K masing-masing K=3, K=5, K=7, K=9 dan K=11. Proses KFold Cross Validation menggunakan skema 10-Fold Cross Validation, dimana dataset akan dibagi menjadi N bagian secara acak. Fold ke-1 adalah Ketika bagian ke-1 menjadi data testing dan sisanya menjadi data training, demikian seterusnya hingga sampai fold ke-10 bagian ke-10. Proses K-Fold Cross Validation ditambahkan pada pengujian dengan tujuan untu meningkatkan performance dari Algoritma KNN.

5) Evaluation

Tahap evaluasi dilakukan untuk melihat performance kinerja klasifikasi yang telah dilakukan pada pemodelan sebelumnya menggunakan confusion matrix untuk melihat nilai accuracy, recall dan precision. Hasil confusion matrix dari Algoritma KNN dengan nilai K=3, K=5, K=7, K=9, dan K=11. Berdasarkan hasil accuracy diatas dapat dilihat nilai accuracy K=5 merupakan accuracy tertinggi dengan menggunakan Algoritma KNN dan 10-Fold Cross Validation mendapatkan nilai accuracy sebesar 81.27%.

6) Deployment

Berdasarkan hasil prediksi tersebut, terdapat 18 Kecamatan dengan kasus demam berdarah yang terjadi relatif tinggi, 7 Kecamatan dengan kasus demam berdarah yang terjadi relatif sedang, dan 5 Kecamatan dengan kasus demam berdarah yang terjadi relatif rendah. Hal tersebut menandakan tingginya kasus Demam Berdarah di Kabupaten Karawang khususnya di 18 Kecamatan tertinggi. Berikut pada Tabel 4.23 merupakan hasil prediksi berdasarkan tiap

Kecamatan. Kemudian untuk lebih jelasnya dilakukan pemetaan dengan tools QGIS dari hasil prediksi yang sudah didapat.

Kesimpulan dari penelitian tersebut dilakukan pengujian dengan melakukan klasifikasi mengenai jumlah kasus demam berdarah di Kabupaten Karawang berdasarkan data pada tahun 2016 – 2020 dengan menggunakan Algoritma K-Neareast Neighbor dan metode 10-Fold Cross Validation yang menghasilkan performance terbaik menggunakan nilai K=5 dan menghasilkan nilai accuracy sebesar 81.27%. Secara keseluruhan, penerapan Algoritma K-Nearest Neighbor dapat memberikan accuracy dan prediksi yang cukup baik. Adapun wilayah yang diprediksi terjadi penyebaran kasus demam berdarah yaitu 18 Kecamatan yang berpotensi tinggi, 7 Kecamatan berpotensi sedang, dan 5 Kecamatan berpotensi rendah. Hasil dari prediksi juga dapat dilihat dari pemetaan yang sudah dibuat.

BAB II
DATASET

Dalam progress ini saya sudah mendapatkan dataset public yang saya dapatkan dalam portal data resmi kota bandung http://data.bandung.go.id/ dengan format csv. Sesuai dengan judul eksperimen saya yaitu adalah memprediksi penyebaran kasus DBD di Kota Bandung maka saya telah mengumpulkan data yang berkaitan dengan eksprimen mengenai DBD dan hal hal yang menunjang atau menyebabkab DBD berkembang sangat cepat.

Dataset yang saya kumpulkan ada 3 dataset yaitu Cuaca Kota Bandung, Populasi Laju Penduduk Kota Bandung, dan Kasus Demam Berdarah di Kota Bandung per Kecamatan.

Jadi setelah saya gabungkan data yang saya dapat terdapat 30 record yaitu terdiri dari Kecamatan, Kepadatan Penduduk, Curah Hujan, Kelembaban, Mobilitas dan Kasus DBD setiap kecamatan.

Table 1 Atribut Dataset

Data	Tipe	Keterangan
Kecamatan	Karakter	Semua Kecamatan di Kota bandung
Kepadatan Penduduk	Numerik	Kepadatan Penduduk tahun 2021
Curah Hujan	Numerik	Curah hujan kota bandung tahun 2021
Kelembaban	Numerik	Temperature udara kota bandung
Mobilitas Penduduk	Numerik	Mobilitas penduduk di kota bandung
Kasus DBD	Kategorikal	Jumlah penduduk yang terkena DBD 2021

BAB III
PROSES EKSPERIMEN

Sesuai dengan Model atau Metode Penelitian yang digunakan yaitu CRISP-DM yang terdiri dari 6 tahapan yaitu Bussiness Understanding, Data Understanding, Data Preparation, Modelling, Evaluation, Deployment. Saat ini perkembangan eksperimen saya:

1. Bussiness Understanding

Tahapan Bussiness Understanding digunakan untuk memahami tujuan yang ingin dicapai dilihat dari perspektif bisnis. beberapa tahapan untuk pemahaman latar belakang dan tujuan diantaranya menentukan tujuan bisnis, menilai situasi dan menentukan tujuan Data Mining.

a. Tujuan Bisnis

Berdasarkan permasalahan yaitu dengan tingginya kasus demam berdarah yang terjadi di beberapa wilayah pada Kota Bandung. Maka dilakukan tujuan untuk memprediksi daerah yang rawan terjangkit demam berdarah di Kota Bandung dengan mengklasifikasikan data menggunakan teknik pemodelan. Sehingga nantinya dapat membantu instansi terkait sebagai referensi dalam mengantisipasi daerah yang rawan demam berdarah di Kota Bandung dengan meninjau hasil dari data yang telah diolah sebelumnya.

b. Melihat Situasi Faktor

Beberapa faktor penyebab demam berdarah yang menjadi informasi dalam penelitian ini. Faktor penyebab demam berdarah yang digunakan pada penelitian ini diantaranya Kepadatan Penduduk, Curah Hujan, Kelembaban, Mobilitas penduduk dan Kasus DBD yang terjadi di Kota Bandung.

c. Tujuan Data Mining

Tujuan Data Mining pada penelitian ini adalah menggali pengetahuan dan pengalaman baru tentang pola (pattern) mengenai daerah penyebaran demam berdarah di Kota Bandung menggunakan algoritma K-Nearest Neighbor. Untuk mendapatkan hasil model terbaik dalam memprediksi, maka menggunakan metode evaluasi dan validasi diantaranya K-Fold Cross Validation untuk

mengevaluasi kinerja dari algoritma KNN dan confusion matrix untuk mengetahui hasil akurasi.

2. Data Understanding

Tahapan Data Understanding digunakan untuk mengenal dan memahami data yang akan diteliti. Proses data understanding ini dilakukan dengan beberapa tahapan untuk pengenalan terhadap data diantaranya proses pengumpulan data dan deskripsi data.

a. Pengumpulan Data

Pada tahap pengumpulan data, data yang digunakan pada penelitian ini bersifat sekunder data yang diambil melalui perantara atau pihak yang telah mengumpulkan data tersebut sebelumnya, dengan kata lain tidak langsung mengambil data sendiri ke lapangan, dimana data diperoleh melalui situs web Pendata Kominfo Kota Bandung (http://data.bandung.go.id/). Terdapat beberapa data yang telah dikumpulkan pada penelitian ini yaitu Kecamatan, Kepadatan Penduduk, Curah Hujan, Kelembaban, Mobilitas dan Kasus DBD setiap kecamatan dimana terdiri dari 30 kecamatan di Kota Bandung.

b. Representasi Pengetahuan

Setelah tahap pengumpulan data, selanjutnya dilakukan pemahaman terhadap data dengan mendeskripsikan masing-masing data. Deskripsi data ini dilakukan untuk memberikan gambaran dari setiap data untuk mengetahui karakteristik data dengan mendiskripsikan tipe dan jenis pada setiap data yang telah dikumpulkan. Berikut merupakan tabel deskripsi dari masing-masing data.

Table 2 Deskripsi Dataset

Data	Tipe	Keterangan
Kecamatan	Karakter/Prediktor	Semua Kecamatan di Kota bandung
Kepadatan Penduduk	Numerik/Prediktor	Kepadatan Penduduk tahun 2021
Curah Hujan	Numerik/Prediktor	Curah hujan kota bandung tahun 2021
Kelembaban	Numerik/Prediktor	Temperature udara kota bandung
Mobilitas Penduduk	Numerik/Prediktor	Mobilitas penduduk di kota bandung

Kasus DBD	Kategorikal/Class/Label	Jumlah penduduk yang terkena DBD 2021

3. Data Preparation

Pada tahap data preparation ini mencakup semua kegiatan dalam menyiapkan dan membangun dataset yang akan diterapkan ke dalam tools pemodelan untuk selanjutnya dilakukan proses Data Mining. Dataset yang digunakan pada penelitian ini sebanyak 150 record data dengan menggunakan 5 parameter sebagai input.

a. Data Cleaning

Pada tahap cleaning, dilakukan pengecekan Missing Value dan dilakukan penanganan dengan mengganti nilai yang missing (tidak diketahui) sesusai data atributnya masing-masing jika ada Missing Value. Namun dalam dataset yang saya miliki tidak memiliki Missing Value.

b. Data Selection

Selanjutnya pada tahap Selection, dilakukan pemilihan data untuk selanjutnya dijadikan atribut yang sesuai dengan kebutuhan pada penelitian. Adapun atribut yang digunakan yaitu data Kepadatan Penduduk, Curah Hujan, Kelembaban, Mobilitas penduduk dan Kasus DBD.

Table 3 Isi Atribut Dataset

Kepadatan Penduduk	Curah Hujan (mm)	Kelembaban (℃)	Mobilitas	Kasus DBD
235	207.6	23.98	3.94	84
188	336.6	23.62	3.00	120
102	290.8	24.05	2.75	139
274	271.4	23.98	2.78	66
200	292.3	24.04	5.97	90
…	…	…	…	…
170	261.8	20.88	3.30	67
194	63.7	20.97	4.37	105
122	63.7	20.5	3.30	100
107	63.7	20.78	1.45	59
141	327.3	20.24	3.04	137

4. Modelling

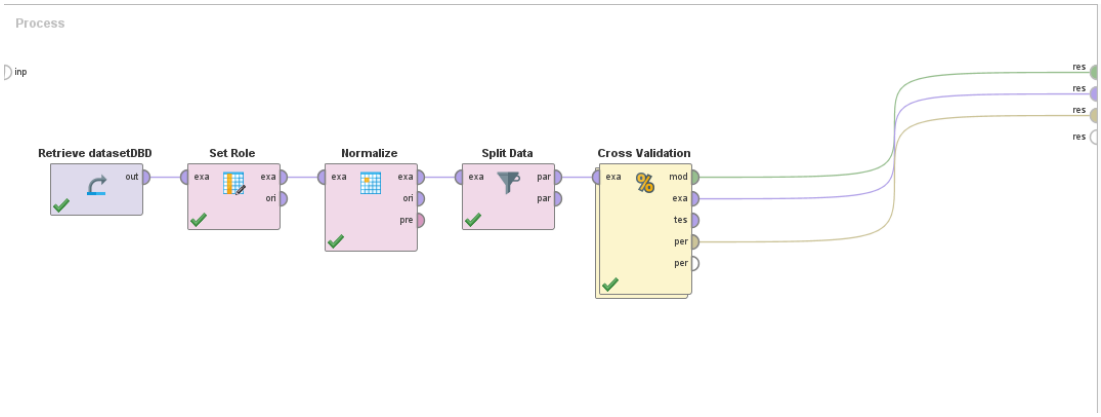
Pada tahap modelling, dilakukan pembagian terhadap dataset secara acak menjadi 2 bagian yaitu data training dan data testing dengan rasio perbandingan 0.8 dan 0.2. Data training digunakan untuk menghasilkan model prediksi menggunakan algoritma KNN dan data testing digunakan untuk melihat performance model prediksi yang dihasilkan. Berikut pada merupakan hasil pembagian terhadap dataset.

Table 4 Proporsi Pembagian Rasio Dataset

Dataset	**Data *Training***			**Data *Testing***		
	Tinggi	Sedang	Rendah	Tinggi	Sedang	Rendah
30	8	14	2	2	4	0

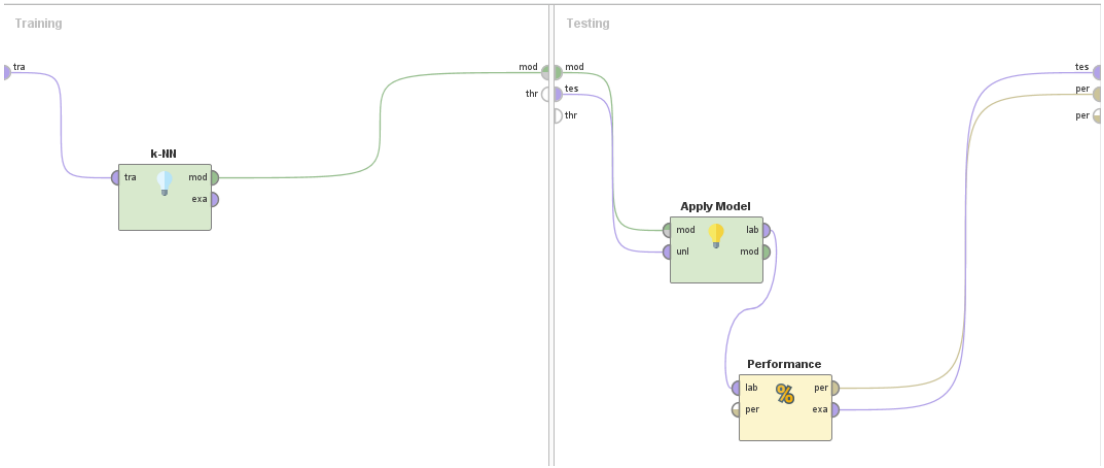
Berdasarkan tabel hasil pembagian dataset diatas, diketahui bahwa data training berjumlah 24 record dengan class tinggi sebanyak 8 record, class sedang 14 record dan class rendah sebanyak 2 record. Sedangkan data testing berjumlah 6 record dengan class tinggi sebanyak 2 record, class sedang 4 record dan class rendah sebanyak 0 record. Sehingga total jumlah class tinggi sebanyak 10 record, class sedang 18 record dan class rendah sebanyak 2 record.

Pengolahan data dilakukan dengan menggunakan tools Rapidminer yang dimana Algoritma KNN dihitung secara otomatis menggunakan tools tersebut. Pada tahap ini dilakukan 5 kali pengujian dengan nilai K masing-masing K=3, K=5, K=7, K=9 dan K=11. Proses K-Fold Cross Validation menggunakan skema 10-Fold Cross Validation, dimana dataset akan dibagi menjadi N bagian secara acak. Fold Proses K-Fold Cross Validation ditambahkan pada pengujian dengan tujuan untuk meningkatkan performance dari Algoritma KNN.



Gambar 1 Gambar Proses Pengujian Data

Berdasarkan gambar diatas terdapat operator Cross Validation yang di dalamnya terdapat subproses. Berikut merupakan rangkaian subproses pada cross validation. Pada subproses training terdapat operator KNN dan pada subroses testing terdapat operator apply model dan performance.



Gambar 2 Subprocess Operator Cross Validation

Terdapat beberapa operator yang digunakan dalam proses pengujian data sebagai berikut:

- Retrieve KasusDBD: Operator ini dapat mengakses informasi yang disimpan di Repositori dan memuatnya ke dalam Proses.
- Set Role: Operator ini untuk memilih salah satu atribut untuk dijadikan class.
- Normalize: Operator ini menormalisasikan nilai Atribut yang dipilih.

- Split Data: Operator ini menghasilkan pembagian data yang diinginkan dari Dataset yang diberikan. Dataset dipartisi menjadi data training dan data testing sesuai dengan ukuran relatif yang ditentukan.
- Cross Validation: Operator untuk evaluasi model.
- KNN: Operator ini menghasilkan model K-Nearest Neighbor yang digunakan untuk klasifikasi data.
- Apply Model: Operator ini menerapkan model pada ExampleSet.
- Performance: Operator ini digunakan untuk evaluasi kinerja statistik tugas klasifikasi. Operator ini memberikan daftar nilai kriteria kinerja tugas klasifikasi, pada penelitian ini dilihat akurasinya.

5. Evaluation

Tahap evaluasi dilakukan untuk melihat performance kinerja klasifikasi yang telah dilakukan pada pemodelan sebelumnya menggunakan confusion matrix untuk melihat nilai accuracy, recall dan precision. Berikut pada Tabel 10 – Tabel 14 merupakan hasil confusion matrix dari Algoritma KNN dengan nilai K=3, K=5, K=7, K=9, dan K=11.

accuracy: 76.67% +/- 25.09% (micro average: 79.17%)

	true Sedang	true Tinggi	true Rendah	class precision
pred. Sedang	13	2	2	76.47%
pred. Tinggi	1	6	0	85.71%
pred. Rendah	0	0	0	0.00%
class recall	92.86%	75.00%	0.00%	

Gambar 3 Hasil Confusion Matrix k = 3

accuracy: 78.33% +/- 23.64% (micro average: 79.17%)

	true Sedang	true Tinggi	true Rendah	class precision
pred. Sedang	14	3	2	73.68%
pred. Tinggi	0	5	0	100.00%
pred. Rendah	0	0	0	0.00%
class recall	100.00%	62.50%	0.00%	

Gambar 4 Hasil Confusion Matrix K = 5

accuracy: 70.00% +/- 26.99% (micro average: 70.83%)

	true Sedang	true Tinggi	true Rendah	class precision
pred. Sedang	14	5	2	66.67%
pred. Tinggi	0	3	0	100.00%
pred. Rendah	0	0	0	0.00%
class recall	100.00%	37.50%	0.00%	

Gambar 5 Hasil Confusion Matrix K = 7

accuracy: 66.67% +/- 29.40% (micro average: 66.67%)

	true Sedang	true Tinggi	true Rendah	class precision
pred. Sedang	14	6	2	63.64%
pred. Tinggi	0	2	0	100.00%
pred. Rendah	0	0	0	0.00%
class recall	100.00%	25.00%	0.00%	

Gambar 6 Hasil Confusion Matrix K = 9

accuracy: 66.67% +/- 29.40% (micro average: 66.67%)

	true Sedang	true Tinggi	true Rendah	class precision
pred. Sedang	14	6	2	63.64%
pred. Tinggi	0	2	0	100.00%
pred. Rendah	0	0	0	0.00%
class recall	100.00%	25.00%	0.00%	

Gambar 7 Hasil Confusion Matrix K = 11

Berdasarkan tabel confusion matrix yang di dapat. Dapat dilihat nilai class recall dan class precision.

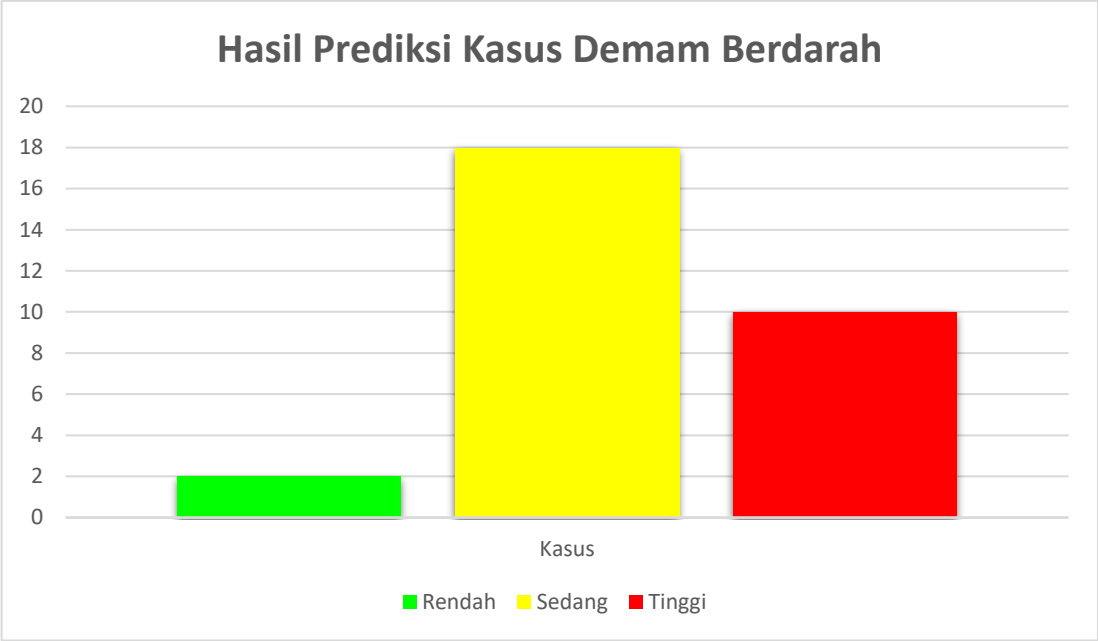
Table 5 Hasil Pengujian Data

KNN	Nilai K				
	K = 3	K = 5	K = 7	K = 9	K = 11
Accuracy	76.67%	78.33%	70.00%	66.67%	66.67%

Berdasarkan hasil accuracy diatas dapat dilihat nilai accuracy K=5 merupakan accuracy tertinggi dengan menggunakan Algoritma KNN dan 10-Fold Cross Validation mendapatkan nilai accuracy sebesar 78.33%.

6. Deployment

Deployment merupakan tahap terakhir dimana hasil penelitian dipresentasikan dalam bentuk yang mudah dipahami mengenai pengetahuan yang diperoleh dalam proses Data Mining. Berdasarkan penelitian yang dilakukan, telah telah menghasilkan suatu informasi dan pengetahuan baru dalam proses Data Mining untuk klasifikasi kasus Demam Berdarah di Kota Bandung. Algoritma K-Nearest Neighbor digunakan untuk mencari nilai akurasi yang menghasilkan prediksi kasus Demam Berdarah pada setiap kecamatan di Kota Bandung.



Gambar 8 Diagram Hasil Prediksi

Berdasarkan hasil prediksi tersebut, terdapat 2 Kecamatan dengan kasus demam berdarah yang terjadi relatif rendah, 18 Kecamatan dengan kasus demam berdarah yang terjadi relatif sedang, dan 10 Kecamatan dengan kasus demam berdarah yang terjadi relatif tinggi. Hal tersebut menandakan penanganan demam berdarah di Kota Bandung masih berjalan dalam hasil prediksi hanya terdapat 10 Kecamatan yang kasusnya relative tinggi.

Kemudian untuk lebih jelasnya dilakukan pemetaan dengan tools QGIS dari hasil prediksi yang sudah didapat.

Gambar 9 Hasil Pemetaan QGIS

KESIMPULAN

Pengujian ini melakukan klasifikasi mengenai jumlah kasus demam berdarah di Kota Bandung berdasarkan data pada tahun 2020 -2021 dengan menggunakan Algoritma K-Neareast Neighbor dan metode 10-Fold Cross Validation yang menghasilkan performance terbaik menggunakan nilai K=5 dan menghasilkan nilai accuracy sebesar 78.33%. Secara keseluruhan, penerapan Algoritma K-Nearest Neighbor dapat memberikan accuracy dan prediksi yang cukup baik. Adapun wilayah yang diprediksi terjadi penyebaran kasus demam berdarah yaitu 10 Kecamatan yang berpotensi tinggi, 18 Kecamatan berpotensi sedang, dan 2 Kecamatan berpotensi rendah. Hasil dari prediksi juga dapat dilihat dari pemetaan yang sudah dibuat.

DAFTAR PUSTAKA

Budiman, Q., Mouton, S., Veenhoff, L., & Boersma, A. (2021). PREDIKSI PENYEBARAN KASUS DEMAM BERDARAH DENGUE(DBD) DENGAN METODE K-NEAREST NEIGHBOR DI KABUPATEN KARAWANG. *Jurnal Inovasi Penelitian*, *1*(0.1101/2021.02.25.432866), 1–15.

Hasanah, M. A., Soim, S., & Handayani, A. S. (2021). Implementasi CRISP-DM Model Menggunakan Metode Decision Tree dengan Algoritma CART untuk Prediksi Curah Hujan Berpotensi Banjir. *Journal of Applied Informatics and Computing*, *5*(2), 103–108. https://doi.org/10.30871/jaic.v5i2.3200