



SEKOLAH TINGGI TEKNOLOGI TERPADU NURUL FIKRI

**PENGGUNAAN ALGORITMA *DECISION TREE* UNTUK PREDIKSI CUACA
DI JAKARTA, BOGOR, DAN TANGERANG MENGGUNAKAN KNIME
*PLATFORM***

TUGAS AKHIR

Raafi Alfarizi

0110221269

PROGRAM STUDI TEKNIK INFORMATIKA

DEPOK

JULI 2025



SEKOLAH TINGGI TEKNOLOGI TERPADU NURUL FIKRI

**PENGGUNAAN ALGORITMA *DECISION TREE* UNTUK PREDIKSI CUACA
DI JAKARTA, BOGOR, DAN TANGERANG MENGGUNAKAN KNIME
*PLATFORM***

TUGAS AKHIR

Diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer

**RAAFI ALFARIZI
0110221269**

**PROGRAM STUDI TEKNIK INFORMATIKA
DEPOK
2025**

HALAMAN PERNYATAAN ORISINALITAS

Skripsi/Tugas Akhir ini adalah hasil karya penulis, dan semua sumber baik yang dikutip maupun dirujuk telah saya nyatakan dengan benar.

**Nama : Raafi Alfarizi
NIM : 0110221269**

Tempat, Depok 10 Juli 2025

Tanda Tangan



A handwritten signature in black ink, appearing to read "Raafi Alfarizi".

Raafi Alfarizi

HALAMAN PENGESAHAN

Skripsi/Tugas Akhir ini diajukan oleh :

Nama : Raafi Alfarizi

NIM : 0110221269

Program Studi : Teknik Informatika

Judul Skripsi : Penggunaan Algoritma Decision Tree Untuk Prediksi Cuaca Di Jakarta, Bogor, Dan Tangerang Menggunakan Knime Platform

Telah berhasil dipertahankan di hadapan Dewan Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Komputer pada Program Studi Teknik Informatika, Sekolah Tinggi Teknologi Terpadu Nurul Fikri

DEWAN PENGUJI

Pembimbing



Dr. Sirojul Munir, S.Si., M.Kom.

Penguji



Bambang Harie Wiyono, S.T., M.M.

Ditetapkan di : Depok

Tanggal : 14 Juli 2025

KATA PENGANTAR

Puji syukur penulis panjatkan kehadirat Allah SWT, karena atas berkat dan rahmat-Nya, penulis dapat menyelesaikan skripsi/Tugas Akhir ini. Penulisan skripsi/Tugas Akhir ini dilakukan dalam rangka memenuhi salah satu syarat untuk mencapai gelar Sarjana komputer Program Studi Teknik Informatika pada Sekolah Tinggi Teknologi Terpadu Nurul Fikri. Penulis menyadari bahwa, tanpa bantuan dan bimbingan dari berbagai pihak, dari masa perkuliahan sampai pada penyusunan skripsi ini, sangatlah sulit bagi penulis untuk menyelesaikan skripsi/tugas akhir ini. Oleh karena itu, penulis mengucapkan terima kasih kepada:

1. Allah SWT.
2. Orang tua dan semua anggota keluarga yang telah memberikan dorongan baik secara moril maupun materil dalam penyelesaian tugas ini.
3. Bapak Dr. Lukman Rosyidi, S.T., M.M., M.T. selaku Ketua Sekolah Tinggi Teknologi Terpadu Nurul Fikri.
4. Ibu Tifani Nabarian, S.Kom., M.T.I. selaku Ketua Program Studi Teknik Informatika Sekolah Tinggi Teknologi Terpadu Nurul Fikri.
5. Bapak Reza Maulana, S.Kom., M.Kom. selaku Dosen Pembimbing Akademik yang telah membimbing penulis selama berkuliah di Sekolah Tinggi Teknologi Terpadu Nurul Fikri.
6. Bapak Dr. Sirojul Munir, S.Si, M.Kom. selaku Dosen Pembimbing Tugas Akhir penulis dalam menyelesaikan penulisan ilmiah ini.
7. Bapak Bambang Harie Wiyono, S.T., M.M. selaku Dosen Penguji Tugas Akhir penulis dalam menyelesaikan penulisan ilmiah ini.
8. Para Dosen di lingkungan Sekolah Tinggi Teknologi Terpadu Nurul Fikri yang telah membimbing penulis dalam menuntut ilmu yang telah diberikan.
9. BMKG Manajer beserta karyawan yang telah meluangkan waktunya untuk memberikan data yang diperlukan bagi penulisan ilmiah ini.

Dalam penulisan ilmiah ini tentu saja masih banyak terdapat kekurangan-kekurangan yang mungkin disebabkan oleh keterbatasan kemampuan dan pengetahuan yang penulis miliki. Walaupun demikian, penulis telah berusaha menyelesaikan penulisan ilmiah ini sebaik mungkin. Oleh karena itu apabila terdapat kekurangan di dalam penulisan ilmiah ini, dengan rendah hati penulis menerima kritik dan saran dari

pembaca.

Akhir kata, penulis berharap Allah SWT berkenan membala segala kebaikan semua pihak yang telah membantu. Semoga skripsi ini membawa manfaat bagi pengembangan ilmu.

Depok, 21 Mei 2025



Raafi Alfarizi

HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI
TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademik Sekolah Tinggi Teknologi Terpadu Nurul Fikri, saya yang bertanda tangan di bawah ini:

Nama : Raafi Alfarizi

NIM 0110221269

Program Studi : Teknik Informatika

Jenis karya : Tugas Akhir

demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada STT-NF Hak Bebas Royalti Noneksklusif (*Non-exclusive Royalty - Free Right*) atas karya ilmiah saya yang berjudul : Algoritma *Decision Tree* Untuk Prediksi Cuaca Di Jakarta, Bogor, Dan Tangerang Menggunakan Knime *Analytics Platform* beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Noneksklusif ini STT-NF berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Depok

Pada tanggal : 21 Mei 2025

Yang Menyatakan



Raafi Alfarizi

ABSTRAK

Nama : Raafi Alfarizi

NIM 0110221269

Program Studi : Teknik Informatika

Judul : Penggunaan Algoritma *Decision Tree* Untuk Prediksi Cuaca Di Jakarta, Bogor, Dan Tangerang Menggunakan Knime *Platform*

Tugas Akhir/Skripsi ini membahas tentang prediksi cuaca atau iklim suatu daerah di Indonesia dengan data yang bersumber dari BMKG yang diterapkan menggunakan pemodelan *machine learning*. Cuaca mengacu pada semua peristiwa yang terjadi di atmosfer di permukaan bumi dalam jangka waktu singkat atau dalam beberapa hari. Beberapa faktor yang mempengaruhi kondisi cuaca yaitu suhu, tekanan udara, kecepatan angin, udara dan curah hujan. Penelitian ini bertujuan untuk memprediksi suhu pada zona cuaca lokal di wilayah Jakarta, Bogor, dan Tangerang. Pentingnya prediksi cuaca karena hal ini dapat membantu menginformasikan sistem peringatan dini sehingga warga bisa mengambil tindakan buat mengurangi bencana sebelum terjadi. Penelitian ini menggunakan metode *classification model decision tree* untuk mengklasifikasikan atau memprediksi suatu nilai target berdasarkan ciri-ciri (fitur) yang ada. Metode ini menggunakan struktur seperti pohon, dimana setiap simpul internal (*node*) mewakili suatu tes atau pemeriksaan pada suatu ciri, setiap cabang mewakili hasil dari tes tersebut, dan simpul daun (*leaf*) mewakili kelas atau kategori hasil prediksi. Hasil *confusion matrix* yang digunakan untuk evaluasi berupa nilai akurasi, *Cohen's kappa*, dan nilai *error* yang dimana akurasi pada model decision tree mencapai 98,187%, kemudian nilai *Cohen's kappa* sebesar 0,935%, dan kesalahan prediksi atau *error* senilai 1,813%. Dari penelitian ini menunjukkan bahwa tingkat akurasi model prediksi cuaca merupakan akurasi yang bagus, serta *cohen's kappa* dari model prediksi cuaca juga merupakan kesepakatan yang sangat baik.

Kata kunci: Prediksi cuaca, Pembelajaran Mesin, *Decision Tree*, BMKG, KNIME *Analytics Platform*

ABSTRACT

Name : Raafi Alfarizi

NIM : 0110221269

Study Program : Information Technology

Title : Decision Tree Algorithm For Weather Prediction In Jakarta, Bogor, And Tangerang Using Knime Analytics Platform

This Final Project/Thesis discusses weather or climate predictions in an area in Indonesia with data sourced from BMKG which is applied using machine learning modeling. Weather refers to all events that occur in the atmosphere on the earth's surface in a short period of time or within a few days. Several factors that affect weather conditions are temperature, air pressure, wind speed, air and rainfall. This study aims to predict temperatures in local weather zones in the Jakarta, Bogor, and Tangerang areas. The importance of weather predictions is because it can help inform early warning systems so that residents can take action to reduce disasters before they occur. This study uses the classification model decision tree method to classify or predict a target value based on existing characteristics (features). This method uses a tree-like structure, where each internal node (node) represents a test or examination of a feature, each branch represents the results of the test, and the leaf node (leaf) represents the class or category of prediction results. The results of the confusion matrix used for evaluation are in the form of accuracy values, Cohen's kappa, and error values where the accuracy of the decision tree model reaches 98.187%, then the Cohen's kappa value is 0.935%, and the prediction error or error is 1.813%. From this study, it shows that the level of accuracy of the weather prediction model is good accuracy, and Cohen's kappa from the weather prediction model is also a very good agreement..

Key words : Weather Forecast, Machine Learning, Decision Tree, BMKG, KNIME Analytics Platform

DAFTAR ISI

HALAMAN PERNYATAAN ORISINALITAS	3
HALAMAN PENGESAHAN	4
KATA PENGANTAR	5
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI	3
TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS.....	3
ABSTRAK	4
ABSTRACT	5
DAFTAR ISI	6
DAFTAR GAMBAR	9
DAFTAR TABEL.....	12
DAFTAR LAMPIRAN	13
BAB I.....	14
PENDAHULUAN	14
1.1 Latar Belakang	14
1.2 Rumusan Masalah.....	15
1.3 Tujuan dan Manfaat Penelitian.....	15
1.4 Batasan Masalah.....	16
1.5 Sistematika Penulisan	16
BAB II	18
KAJIAN LITERATUR.....	18
2.1 Tinjauan Pustaka	18
2.1.1 Cuaca dan Iklim	18
2.1.2 Data Mining	18
2.1.3 Machine learning	21

2.1.4 Metode Pengujian	22
1. Confusion Matrix.....	22
2. Cohen's kappa	22
3. Accuracy.....	23
2.1.5 KNIME Analytics Platform.....	24
2.2 Penelitian Terkait	27
2.3 Posisi Penelitian	30
BAB III.....	32
ANALISIS DAN PERANCANGAN SISTEM	32
3.1 Tahapan Penelitian.....	32
3.1.1 Studi Pendahuluan.....	32
3.1.2 Analisis Permasalahan.....	33
3.1.3 Perancangan Sistem.....	33
3.1.4 Implementasi Sistem.....	33
3.1.5 Evaluasi.....	34
3.1.6 Kesimpulan	34
3.2 Jenis Penelitian	35
3.3 Metode Analisis Data	35
3.4 Metode Pengumpulan Data	35
3.5 Metode Pengujian.....	36
3.6 Evaluasi Hasil Pengujian	36
3.7 Lingkungan Pengembangan	36
BAB IV	38
IMPLEMENTASI DAN EVALUASI	38
4.1 Business Understanding.....	38
4.2 Data Understanding	38
4.3 Data Preparation	40

4.3.1 Data Transformation.....	41
4.3.2 Data Integration.....	55
4.3.3 Data Cleaning.....	57
4.3.4 Feature Selection.....	66
4.4 Modelling.....	71
4.5 Evaluation.....	81
BAB V.....	87
KESIMPULAN DAN SARAN.....	87
5.1 Kesimpulan	87
5.2 Saran.....	87
DAFTAR REFERENSI.....	89
LAMPIRAN.....	92

DAFTAR GAMBAR

Gambar 2.1 Alur kerja untuk melatih data.....	25
Gambar 2.2 Alur kerja untuk visualisasi data	25
Gambar 2.3 Filtering data.....	26
Gambar 2.4 Fungsi logika	26
Gambar 2.5 Data Transformation.....	27
Gambar 2.6 Tahapan Penelitian	32
Gambar 4. 1 Tampilan node csv reader untuk membaca data climate_data.csv.....	40
Gambar 4. 2 Tampilan konfigurasi node 2 untuk membaca data station_detail.csv	40
Gambar 4. 3 Data Transformation 1	41
Gambar 4. 4 Data Transformation 2	42
Gambar 4. 5 Alur kerja data transformation 2	42
Gambar 4. 6 Alur kerja untuk filtering data	43
Gambar 4. 7 Alur kerja fungsi logika dan perubahan tipe data untuk data transformation.....	43
Gambar 4. 8 Data transformation 3.....	44
Gambar 4. 9 Konfigurasi node column renamer untuk merubah nama kolom.....	44
Gambar 4. 10 Kolom selanjutnya yang akan di rename	45
Gambar 4. 11 Tampilan tabel data setelah kolom di rename	45
Gambar 4. 12 Konfigurasi node row filter untuk exclude baris 1-48 yang tidak diperlukan	46
Gambar 4. 13 Konfigurasi node row id untuk mengubah kembali baris data station_detail.csv yang sebelumnya bukan dari baris 1	46
Gambar 4. 14 Konfigurasi untuk mengurutkan data climate_data.csv	47
Gambar 4. 15 Exclude baris 1-9	47
Gambar 4. 16 Exclude baris 1 - 50.954	48
Gambar 4. 17 Exclude baris 1 - 105.851	48
Gambar 4. 25 Exclude baris 188.906 - 372.150	49

Gambar 4. 27 Exclude baris 182.411 - 185.298	49
Gambar 4. 29 Kode program/sintax untuk konfigurasi node rule engine	50
Gambar 4. 30 Tampilan isi kolom sebelum di konfigurasi	50
Gambar 4. 31 Tampilan isi baris pada kolom setelah dikonfigurasi	51
Gambar 4. 32 Konfigurasi node number to string.....	51
Gambar 4. 33 Kolom data sebelum di konfigurasi number to string.....	52
Gambar 4. 34 Kolom data sesudah di konfigurasi menggunakan number to string	52
Gambar 4. 35 Konfigurasi string manipulation.....	53
Gambar 4. 37 Tampilan kolom data setelah dikonfigurasi.....	54
Gambar 4. 38 Konfigurasi node sorter untuk mengurutkan baris dari yang terlama ke terbaru.....	54
Gambar 4. 39 Tampilan kolom data setelah dikonfigurasi yang pada gambar sebelumnya belum berurutan	54
Gambar 4. 40 Konfigurasi row id untuk mengatur nomor baris secara berurutan.....	55
Gambar 4. 41 Node joiner untuk tahap data integration	56
Gambar 4. 42 Konfigurasi node joiner tab joiner settings.....	56
Gambar 4. 43 Konfigurasi node joiner tab column selection	57
Gambar 4. 44 Tampilan gabungan 2 data	57
Gambar 4. 45 Tahapan data cleaning	58
Gambar 4. 46 Isi dari metanode data cleaning.....	58
Gambar 4. 47 Konfigurasi untuk menghapus baris 1 - 88.984.....	59
Gambar 4. 48 Tampilan tabel data setelah di konfigurasi node row filter.....	59
Gambar 4. 49 Tampilan tabel data sebelum dikonfigurasi.....	60
Gambar 4. 50 Konfigurasi untuk menetapkan kembali ID baris	60
Gambar 4. 51 Tampilan tabel data setelah konfigurasi row id	61
Gambar 4. 52 Menghapus baris 7.841 - 11.760.....	61
Gambar 4. 53 Tampilan tabel data setelah dikonfigurasi menggunakan node row filter.....	62
Gambar 4. 54 Menghapus baris 38.175 - 157.133.....	62
Gambar 4. 55 tampilan tabel data setelah dikonfigurasi menggunakan node row id.....	63

Gambar 4. 56 Konfigurasi untuk menetapkan kembali ID baris	63
Gambar 4. 57 Konfigurasi untuk menghapus baris 34.255 - 157.133	64
Gambar 4. 58 Konfigurasi untuk menghilangkan nilai kosong menggunakan node missing value.....	64
Gambar 4. 59 Konfigurasi node missing value tab column settings.....	65
Gambar 4. 60 Tabel Data Setelah Dikonfigurasikan Menggunakan Node Missing Value	65
Gambar 4. 61 Konfigurasi untuk menetapkan kembali ID baris	66
Gambar 4. 62 Metanode tahap feature selection.....	67
Gambar 4. 63 Konfigurasi node rank correlation.....	67
Gambar 4. 64 Hasil correlation matrix	68

DAFTAR TABEL

Tabel 2.1 Tabel Penelitian Terkait.....	27
Table 2.2 Posisi Penelitian	30
Tabel 3.5 1 Tabel Confusion Matrix	37
Tabel 3.5 2 Tabel Cohen's Kappa	38
Tabel 4. 1 Jenis kolom dari sumber data.....	39
Tabel 4. 2 Alasan pemilihan fitur	71

DAFTAR LAMPIRAN

Lampiran A. 1 climate_data.csv	92
Lampiran A. 2 station_detail.csv	92
Lampiran B. 1 Aplikasi untuk alat bantu	93
Lampiran B. 2 Tampilan isi aplikasi beserta alur kerja	93

BAB I

PENDAHULUAN

1.1 Latar Belakang

Saat ini teknologi informasi dan komunikasi (ICT) berkembang dengan pesat yang berdampak pada kemudahan bagi masyarakat dalam mendukung kelancaran aktifitas keseharian. Berbagai informasi dapat diakses oleh masyarakat dengan cepat dan akurat melalui perangkat ICT seperti mengakses data perubahan cuaca atau iklim untuk keperluan aktifitas berbagai bidang kehidupan, seperti bidang pertanian[1], penerbangan dan pelayaran dan bidang lainnya[2].

Indonesia mengalami banyak bencana setiap tahunnya. Contohnya termasuk curah hujan ekstrim dan perubahan suhu yang dramatis. Selain itu, dari segi hidroklimat, Indonesia juga terkena dampak fenomena ENSO (*El Niño Southern Oscillation*) dan La Niña yang menyebabkan banjir, tanah longsor, kekeringan, angin puting beliung, dan suhu rendah [3].

Saat itu, iklim Indonesia terkadang tidak stabil. Bahkan, ketika cuaca di suatu daerah terlihat cerah, bisa saja tiba-tiba berubah menjadi hujan atau bahkan badai. Sulitnya memprediksi kondisi cuaca disebabkan oleh perubahan iklim yang tidak dapat diprediksi. Menurut Badan Meteorologi, Klimatologi, dan Geofisika (BMKG), suhu rata-rata bulan Juli di Indonesia selama tahun 2010 adalah 26,39°C. Suhu rata-rata pada bulan Februari 2020 adalah 27,22°C, dan suhu rata-rata meningkat secara tidak normal sebesar 0,83°C[4].

Cuaca mengacu pada semua peristiwa yang terjadi di atmosfer di permukaan bumi dalam jangka waktu singkat atau dalam beberapa hari[5]. Beberapa faktor yang mempengaruhi kondisi cuaca yaitu suhu, tekanan udara, kecepatan angin, udara dan curah hujan. Kepadatan penduduk dan pesatnya pembangunan infrastruktur di wilayah Jakarta dan sekitarnya (Jakarta, Bogor, dan Tangerang) menyebabkan perubahan cuaca perkotaan yang signifikan[5]. Penelitian ini bertujuan untuk memprediksi suhu pada zona cuaca lokal di wilayah Jakarta, Bogor, dan Tangerang. Pentingnya prediksi cuaca karena hal ini dapat membantu menginformasikan sistem peringatan dini sehingga

warga bisa mengambil tindakan buat mengurangi bencana sebelum terjadi.

Pemodelan *machine learning* banyak diterapkan dalam memprediksi cuaca atau iklim suatu daerah di Indonesia [6], dengan data yang bersumber dari BMKG. Beberapa model *machine learning* yang diterapkan antara lain: *Support Vector Machine* (SVM)[5], *Neural Network*[7], dan *Decision Tree*[8]. *Random Forest*[6] dengan alat bantu aplikasi pendukung seperti aplikasi Weka[8], dan pustaka program *python sklearn*[6].

Berdasarkan latar belakang diatas penelitian ini akan menerapkan *machine learning* untuk prediksi cuaca di Jakarta, Bogor, dan Tangerang dengan judul penelitian “Penggunaan Algoritma *Decision Tree* Untuk Prediksi Cuaca Di Jakarta, Bogor, dan Tangerang Menggunakan KNIME Analytics Platform”.

1.2 Rumusan Masalah

1. Bagaimana memprediksi cuaca di Jakarta, Bogor, dan Tangerang dengan algoritma *Decision Tree* menggunakan KNIME *analytics platform*?
2. Bagaimana tingkat akurasi dan *cohen's kappa* model prediksi cuaca yang diterapkan menggunakan model machine learning dengan *Platform KNIME*?

1.3 Tujuan dan Manfaat Penelitian

Adapun tujuan dari penelitian adalah:

1. Memprediksi Cuaca Di Jakarta, Bogor, dan Tangerang dengan algoritma *Decision Tree* menggunakan KNIME *Analytics Platform*.
2. Mendapatkan tingkat akurasi model prediksi cuaca yang diterapkan menggunakan *Platform KNIME*.

Adapun manfaat penelitian ini adalah:

1. Memperoleh gambaran suhu homogen-rata pada bentuk *decision tree*

berdasarkan data cuaca dari tahun 2010 sampai 2020 di wilayah Jakarta, Bogor, dan Tangerang. yang akan terjadi ini berguna bagi penulis, para pakar meteorologi, pemerintah daerah, dan instansi terkait yang membutuhkan referensi prediksi suhu buat analisis iklim serta perencanaan kebijakan.

2. Menilai tingkat akurasi serta *Cohen's Kappa* dari data suhu rata-rata selama periode 2010–2020 pada Jakarta, Bogor, dan Tangerang. Hal ini berguna bagi data scientist, dan pengembang sistem prediksi cuaca yang ingin menguji sejauh mana keandalan model suhu pada mendukung pengambilan keputusan.
3. Menyajikan data suhu rata-rata dalam bentuk visualisasi berdasarkan variabel yang dipilih. Visualisasi ini mampu dimanfaatkan oleh akademisi, analis data, juga warga umum yang ingin memahami pola perubahan suhu menggunakan cara yang lebih intuitif dan praktis dipahami.

1.4 Batasan Masalah

1. Dataset cuaca yang digunakan bersumber dari BMKG dari tahun 2010-2020.
2. Analisis data menggunakan *platform KNIME* versi 5.2.4
3. Algoritma pada penelitian ini hanya menggunakan *Decision tree*

1.5 Sistematika Penulisan

Sistematika penulisan dibagi menjadi beberapa bab yang akan penulis bahas antara lain:

Bab I: Pendahuluan

Berisi tentang Latar belakang mengenai peran penting dalam memprediksi cuaca dengan *machine learning* menggunakan algoritma *decision tree*. Setelah itu, terdapat rumusan masalah yang menjadi fokus kajian, disertai dengan tujuan yang ingin diraih serta manfaat yang diinginkan dari penelitian prediksi cuaca di Jakarta, Bogor, dan Tangerang. Supaya penelitian lebih terarah, batasan masalah dijelaskan secara jelas, serta disajikan sistematika penulisan untuk membantu pembaca memahami alur pembahasan

dalam penelitian ini.

Bab II: Kajian Literatur

Pada bab kajian literatur ini berisi tentang tinjauan pustaka yang meliputi cuaca dan iklim, *data mining*, *machine learning*, metode pengujian, dan alat pengembangan seperti KNIME *Analytics Platform*. Kemudian terdapat penelitian terkait yang dimana menafsirkan penelitian-penelitian sebelumnya yang berhubungan dengan prediksi cuaca menggunakan *machine learning* algoritma *decision tree* meliputi tabel penelitian serta penjelasan terkait penelitian yang diambil, dan posisi penelitian.

Bab III: Analisis dan Perancangan Sistem

Pada bab ini berisi jenis penelitian yang dimana penelitian ini menggunakan Research and Development yang digunakan oleh penulis dan bagaimana penelitian dilakukan. Metode yang digunakan bertujuan untuk Menerapkan Algoritma Decision Tree Untuk Prediksi Cuaca Di Jakarta, Bogor, dan Tangerang Menggunakan KNIME Analytics Platform, dst. Kemudian pada bab ini juga berisi metode pengumpulan data yang dilakukan melalui studi pustaka dengan mengumpulkan data artikel-artikel penelitian terkait penerapan machine learning dan data mining dalam topik untuk memprediksi cuaca atau iklim dunia yang diperoleh melalui BMKG yang di share melalui platform aplikasi online kaggle.com. Dan pada bab ini berisi metode pengujian berdasarkan umpan balik pengguna atau user, yang dilakukan untuk pengujian dengan menggunakan node Scorer yang dimana terdapat Confusion Matrix.

BAB IV : Implementasi Dan Evaluasi

Bab ini berfokus pada implementasi *machine learning* yang dikembangkan dan bagaimana sistem dirancang, hasil uji dan hasil evaluasi. Sistem ini dirancang menggunakan alat pengembangan seperti KNIME Analytics Platform yang digunakan untuk menerapkan machine learning untuk menganalisis cuaca. Bagian ini penting karena menyajikan hasil sebenarnya dari penelitian yang dilakukan.

BAB V : Kesimpulan Dan Saran

Bab ini menyajikan kesimpulan dan saran bertujuan untuk mengumpulkan hasil dari

seluruh penelitian yang telah dikerjakan, serta bertujuan memberikan masukan untuk pengembangan lebih lanjut.

BAB II

KAJIAN LITERATUR

2.1 Tinjauan Pustaka

2.1.1 Cuaca dan Iklim

Cuaca artinya rangkuman info kondisi cuaca harian hingga mingguan, sedangkan prakiraan iklim umumnya artinya unsur – unsur iklim yang umumnya untuk daerah Indonesia adalah prakiraan hujan bulanan atau prakiraan hujan yang berlangsung pada satu informasi modern. Dengan demikian jenis prakiraan cuaca serta iklim dibedakan pada kurun ketika serta jenis unsur yang diprakirakan, dimana prakiraan cuaca lebih banyak menjelaskan hampir seluruh unsur cuaca serta prakiraan iklim umumnya berkisar pada kuantitas curah hujan dan awal musim[8].

Oleh karena itu, sangat penting untuk memahami hubungan antara cuaca dan iklim, seperti yang dijelaskan di atas, terutama dalam konteks adaptasi terhadap perubahan iklim. Perubahan iklim di wilayah perkotaan menantang pada aktivitas dan kebijakan pemerintah daerah dan pusat yang mendukung berbagai sektor kuat, seperti kebijakan pembangunan di sekitar ibukota Jakarta untuk beberapa kota satelit, yaitu Bogor, dan Tangerang. Pemahaman akan perubahan iklim seperti curah hujan yang berlebih akan mengakibatkan bencana banjir menjadi panduan bagi pemerintahan daerah.

Penelitian yang dilakukan di Jakarta, Bogor, dan Tangerang menunjukkan bahwa penggunaan model cuaca dan iklim yang akurat[9] dapat membantu memprediksi kejadian ekstrem, seperti banjir atau kekeringan, serta mengurangi dampak negatifnya terhadap masyarakat. Dengan memahami hubungan ini, pemerintah dapat mempersiapkan diri dan mengambil langkah-langkah yang diperlukan untuk melindungi masyarakat[10].

2.1.2 Data Mining

Data mining merupakan kombinasi bidang ilmu komputer yang berfokus pada proses

penemuan pola baru dari kumpulan data yang sangat besar. Metode yang digunakan dalam data mining merupakan hasil kombinasi disiplin ilmu seperti kecerdasan buatan, pembelajaran mesin, statistika, dan sistem basis data.[11]. Berdasarkan definisi sebelumnya, dapat disimpulkan bahwa data mining adalah proses analisis terhadap kumpulan data atau dataset untuk menemukan hubungan dan pola-pola yang dapat dipahami serta bermanfaat bagi pemilik data[12].

CRISP-DM (*Cross Industry Standard for Data mining*) adalah secara defacto merupakan standar dan proses model yang diterapkan dalam industri ketika mengaplikasikan projek data mining[13]. Beberapa tahapan Crisp-DM adalah sebagai berikut:

1. Business Understanding

Pertama-tama adalah melakukan pemahaman bisnis. *Business Understanding* adalah langkah awal yang penting dalam sebuah penelitian bisnis. Proses ini melibatkan penetapan tujuan yang ingin dicapai, pemahaman tentang situasi dan kondisi yang ada saat penelitian dilakukan, serta penentuan fokus penelitian yang relevan untuk memecahkan masalah melalui analisis data. Dengan memahami aspek-aspek ini, kita dapat mengarahkan upaya penelitian untuk mencapai hasil yang bermanfaat bagi bisnis[14].

2. Data Understanding

Tahapan ini adalah fase analisis dalam proses penelitian, di mana kita menyusun ringkasan dan mengidentifikasi masalah yang ada dalam data. Pada tahap ini, perhatian yang cermat sangat diperlukan agar tidak mengganggu proses pemodelan data di kemudian hari.

Di sini, kita melakukan ringkasan data yang penting untuk mengonfirmasi sebetulnya data yang digunakan sesuai dan dapat diandalkan untuk langkah selanjutnya. Beberapa kegiatan yang dilakukan dalam tahap ini meliputi pengumpulan data yaitu mengumpulkan informasi yang diperlukan dari berbagai sumber, analisis data untuk menemukan pola atau wawasan baru yang dapat membantu dalam pengambilan keputusan, dan evaluasi kualitas data untuk memeriksa dan menilai kualitas data untuk

memastikan bahwa data yang digunakan akurat dan relevan[15].

3. Data Preparation/Preprocessing Data

Pada tahap ini, kita akan menyusun dataset akhir dari data mentah yang telah tersedia. Proses ini melibatkan beberapa langkah penting antara lain pembersihan data atau data cleaning yaitu menghapus atau memperbaiki data yang tidak akurat, duplikat, atau tidak lengkap agar dataset menjadi lebih berkualitas. Kemudian pada data selection yaitu memilih data yang relevan dan penting untuk analisis, sehingga kita hanya menggunakan informasi yang diperlukan, menyusun dan mengorganisir data dalam bentuk yang terstruktur, termasuk menentukan atribut atau fitur yang akan dianalisis, dan *data transformation* yang mengubah data ke dalam format yang sesuai agar dapat digunakan dalam tahap pemodelan selanjutnya[16].

4. Modelling

Modelling yaitu tahap implementasi algoritma yang akan digunakan untuk melakukan pencarian, mempartisi data, dan pelatihan dicadangkan untuk pengujian 20%. 80% dari set data, identifikasi, serta menghasilkan pola yang akan digunakan pada data penelitian berdasarkan jenis data yang dipilih[17].

5. Evaluation

Dalam mengevaluasi model penelitian ini, digunakan 3 metode analisis diantaranya: akurasi, *kappa cohen*[18], dan *Confusion Matrix*. Ketiga metode ini dipilih untuk memberikan evaluasi komprehensif terhadap kinerja model yang dikembangkan. Hal ini penting karena setiap metode memberikan informasi berbeda tentang performa model. Indikator-indikator ini dihitung menggunakan:

$$\text{Akurasi} = \frac{TP + TN}{TP + FP + FN + TN}$$
$$kappa = \frac{p_0 - p_e}{1 - p_e}$$

Confusion Matrix secara spesifik memetakan hasil prediksi model dengan nilai sebenarnya. Dalam *Confusion Matrix*, kita dapat menemukan:

Positif Benar atau *True Positive* (TP): Bilangan ramalan positif benar

Negatif Benar *True Negative* (TN): Bilangan ramalan negatif negatif

Positif Salah *False Positive* (FP): Bilangan ramalan positif salah

Negatif Salah *False Negative* (FN): Bilangan ramalan negatif salah [19]

2.1.3 *Machine learning*

Machine learning atau pembelajaran mesin adalah cabang kecerdasan buatan yang memungkinkan komputer belajar secara otomatis dan meningkatkan kinerjanya tanpa pemrograman eksplisit. Dengan kata lain, pembelajaran mesin menggunakan algoritme untuk menemukan pola dalam data guna membuat prediksi dan keputusan yang akurat.

Dengan menggunakan teknik *machine learning*, prakiraan cuaca dapat dilakukan secara lebih efisien dan akurat, sehingga membawa manfaat yang sangat besar bagi masyarakat[6]. Penerapan *machine learning* dapat membantu di berbagai bidang seperti bidang pertanian dan data penginderaan jarak jauh[20]. Berikut ini adalah algoritma *machine learning* yang digunakan yaitu *Decision Tree*. *Decision Tree* adalah sebuah metode untuk mengklasifikasikan atau memprediksi suatu nilai target berdasarkan ciri-ciri (fitur) yang ada. Metode ini menggunakan struktur seperti pohon, dimana setiap simpul internal (*node*) mewakili suatu tes atau pemeriksaan pada suatu ciri, setiap cabang mewakili hasil dari tes tersebut, dan simpul daun (*leaf*) mewakili kelas atau kategori hasil prediksi.

Decision Tree bekerja dengan cara membuat keputusan bertingkat, mirip dengan pohon keputusan. Setiap simpul di dalam pohon tersebut mewakili suatu tes atau pemeriksaan terhadap suatu ciri, dan berdasarkan hasil tes tersebut, akan mengarahkan ke cabang-cabang yang lain. Proses ini terus berlanjut hingga akhirnya sampai ke simpul daun yang mewakili hasil prediksi atau klasifikasi akhir. Metode ini digunakan untuk menganalisis dan membandingkan probabilitas dari simpul-simpul dan cabang-cabang yang berbeda, sehingga bisa dievaluasi dan diambil keputusan yang terbaik[8].

Decision Tree atau pohon keputusan disertakan dalam metode pembelajaran terawasi dan digunakan untuk mengklasifikasikan data dalam kumpulan data saat

mengambil keputusan [6].

2.1.4 Metode Pengujian

Adapun alat pengujian dari metode prediksi dari *Decision Tree* adalah:

1. *Confusion Matrix*

Confusion Matrix adalah alat yang digunakan untuk mengevaluasi kinerja model klasifikasi. Ini merupakan tabel perbandingan prediksi model dengan nilai sebenarnya (kondisi nyata). Dengan menggunakan matriks konfusi, kita dapat menghitung berbagai metrik penting seperti akurasi, presisi, perolehan, dan skor F1, yang dapat memberi kita gambaran tentang seberapa baik model klasifikasi kita bekerja.

Jadi, *Confusion Matrix* adalah cara yang sangat berguna untuk mengukur dan memahami kinerja model *machine learning*, terutama dalam tugas-tugas klasifikasi. Ini membantu kita mengidentifikasi kekuatan dan kelemahan model, sehingga kita dapat terus memperbaikinya untuk mendapatkan hasil yang lebih akurat[21].

2. *Cohen's kappa*

Cohen's kappa adalah alat penting untuk mengukur seberapa baik model klasifikasi kita bekerja. Ini menunjukkan seberapa akurat prediksi model kita dibandingkan dengan hasil sebenarnya, dengan memperhitungkan kemungkinan kebetulan.

Nilai kappa berkisar dari -1 hingga 1. Jika nilainya 1, itu berarti model kita memprediksi dengan sempurna. Jika nilainya 0, itu berarti model hanya menebak dengan asal-asalan. Dan jika nilainya negatif, itu berarti model bahkan lebih buruk daripada menebak secara acak.

Semakin tinggi nilai kappa, semakin dapat diandalkan model klasifikasi kita untuk membuat prediksi yang akurat pada data baru yang belum pernah dilihat sebelumnya. Ini adalah metrik yang sangat berguna untuk mengevaluasi dan

memperbaiki kinerja model *machine learning*. Adapun rumus *Cohen's kappa* adalah sebagai berikut:

$$\kappa = \frac{p_o - p_e}{1 - p_e}$$

Dalam perhitungan *Cohen's kappa*, ada dua komponen utama:

1. *Observed agreement (Po)*: Ini adalah persentase kesepakatan antara hasil prediksi model dan nilai sebenarnya pada dataset. Ini menunjukkan seberapa sering model memprediksi dengan benar.
2. *Expected agreement (Pe)*: Ini adalah persentase kesepakatan yang diharapkan terjadi secara acak, tanpa model apapun. Ini menunjukkan kemungkinan model menebak dengan benar hanya karena kebetulan.

Cohen's kappa kemudian dihitung dengan membandingkan *observed agreement (Po)* dan *expected agreement (Pe)*. Semakin besar perbedaan antara *Po* dan *Pe*, semakin baik performansi model klasifikasi.

Jadi, *observed agreement* mencerminkan akurasi aktual model, sementara *expected agreement* mencerminkan akurasi yang diharapkan terjadi secara acak. Perbandingan antara keduanya memberikan ukuran seberapa baik model kita bekerja[22].

3. Accuracy

Akurasi adalah salah satu metrik utama untuk mengevaluasi kinerja model klasifikasi. Akurasi mengukur seberapa sering model memprediksi kelas dengan benar, baik kelas positif maupun negatif. Semakin tinggi nilai akurasinya, semakin baik model dalam membuat prediksi yang akurat.

Namun, akurasi saja tidak selalu cukup, terutama jika dataset memiliki ketidakseimbangan kelas (*class imbalance*). Dalam kasus seperti itu, akurasi bisa menjadi metrik yang menyesatkan. Oleh karena itu, perlu mempertimbangkan metrik evaluasi lain seperti presisi, recall, dan skor F1 secara bersamaan dengan akurasi. Adapun rumus dari akurasi yaitu sebagai berikut:

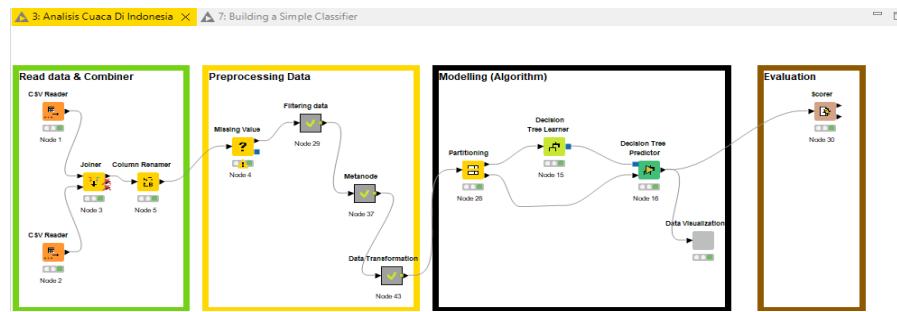
$$Akurasi = \frac{TP+TN}{TP+FP+FN+TN} [23]$$

2.1.5 KNIME Analytics Platform

KNIME adalah perangkat lunak untuk membangun dan memproduksi ilmu data menggunakan lingkungan yang mudah dan intuitif yang memungkinkan setiap pemangku kepentingan dalam proses ilmu data untuk fokus pada apa yang paling mereka kuasai. Ini adalah perangkat lunak ilmu data sumber terbuka yang memfasilitasi implementasi alat penambangan data seperti visualisasi data, algoritma pembelajaran mesin, ekstraksi kering asosiasi. Ini terdiri dari alur kerja *node*. KNIME memiliki banyak anggota di seluruh dunia.

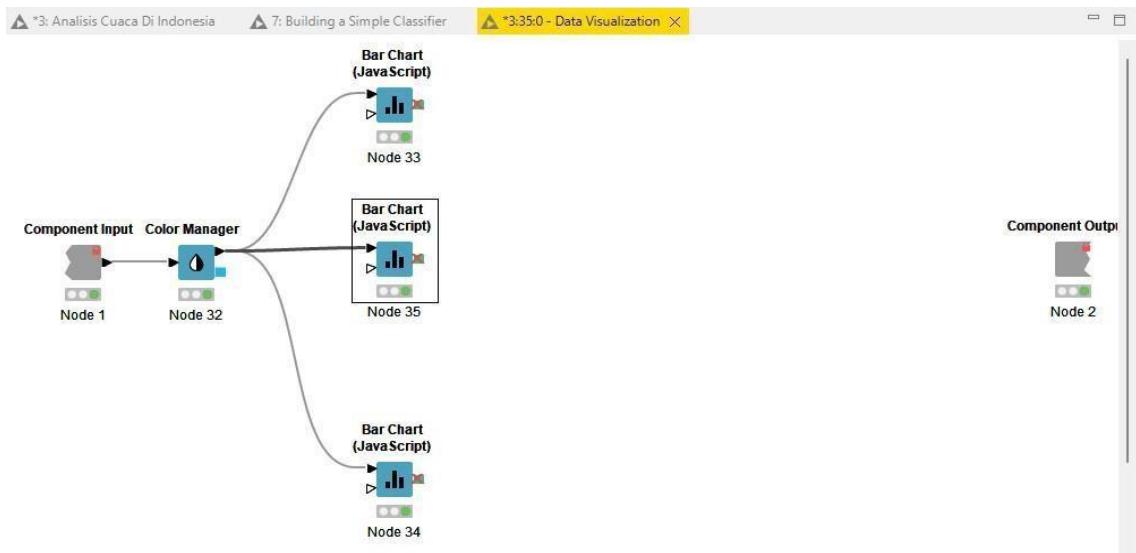
Platform KNIME adalah perangkat lunak ilmu data *Open Source* yang memfasilitasi implementasi *machine learning* dan alat penambangan data, asosiasi data beserta visualisasi data. KNIME adalah *platform Open Source* yang memiliki banyak anggota di seluruh dunia. Para pengguna ini berbagi pekerjaan mereka di situs resmi KNIME, yang memungkinkan pengguna lain untuk mengaksesnya. Pengembangan KNIME dimulai tahun Januari 2004 oleh tim *software engineer* dari Universitas Konstanz sebagai salah satu produk kekayaan intelektual kampus tersebut. Ketua tim engineer yaitu Mr. Michael Berthold dari perusahaan *Silicon Valley* yang menyediakan aplikasi untuk perusahaan farmasi. Selanjutnya pada tahun 2006 tercipta KNIME versi 1 yang dirilis untuk beberapa perusahaan farmasi. Kemudian setelah itu digunakan secara luas sejak tahun 2012 digunakan oleh 15.000 pengguna.

Para pengguna ini membagikan pekerjaan mereka di situs resmi KNIME, yang memungkinkan pengguna lain untuk mengaksesnya. Selanjutnya adalah *Microsoft excel* (File csv) yang dimasukkan ke aplikasi KNIME, sehingga tampil sebuah *node* di KNIME yang bernama *CSV Reader* untuk alat pembaca file tersebut. Adapun *node* yang digunakan pada KNIME yaitu *CSV Reader*, *Statistics*, *Column Renamer*, *Missing Value*, *Partitioning*, *Decision Tree Learner* dan *Predictor*, *Data Visualization*, dan *Scorer*[24]. Gambar 2.1 berikut ini adalah alur kerja/*workflow* dari beberapa simpul/*node* yang dipasangkan untuk melatih data:



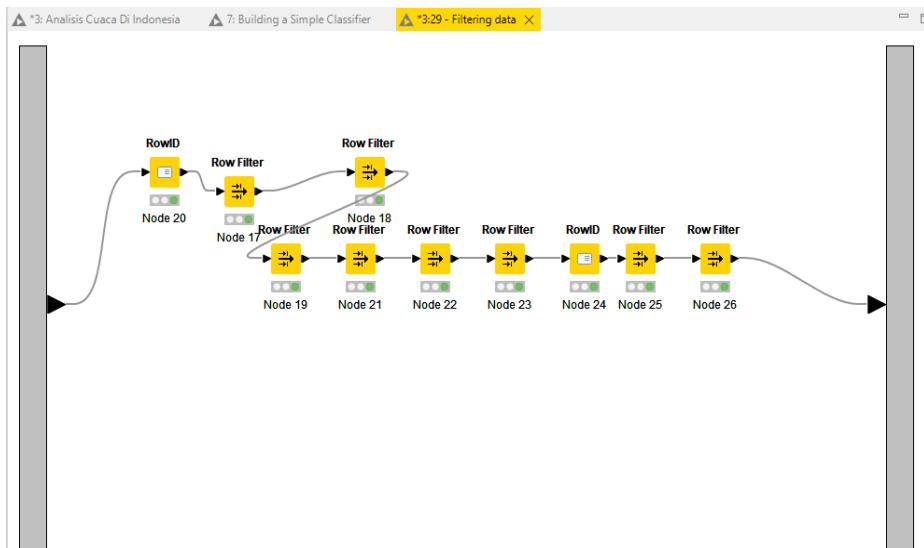
Gambar 2.1 Alur kerja untuk melatih data

Alur kerja/*Workflow* selanjutnya dalam *Component*, Terdapat simpul-simpul/*node* yang dipasangkan sebagai visualisasi data pada Gambar 2.2 dibawah ini:



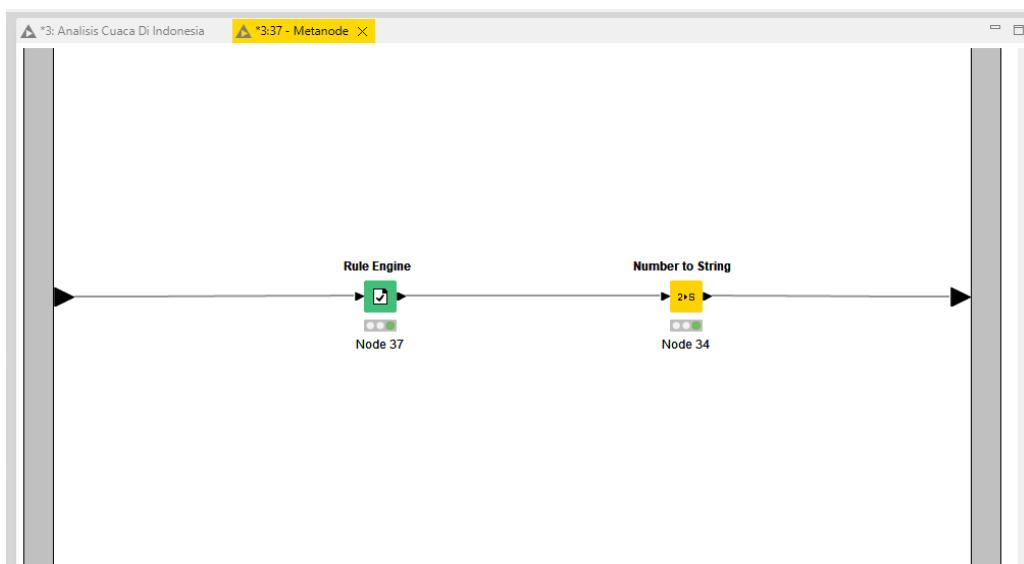
Gambar 2.2 Alur kerja untuk visualisasi data

Alur kerja/Workflow selanjutnya dalam *metanode* yang terpasang setelah *node Normalizer*, Terdapat simpul-simpul/*node* yang dipasangkan sebagai *filtering* data pada Gambar 2.3 dibawah ini:



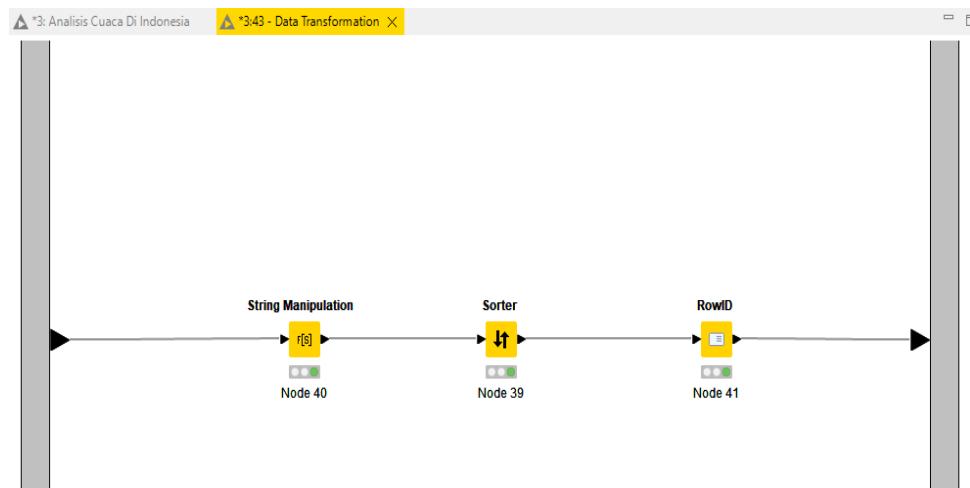
Gambar 2.3 Filtering data

Alur kerja/Workflow selanjutnya dalam *metanode*, Terdapat simpul-simpul/*node* yang dipasangkan sebagai mengubah tipe data menggunakan fungsi logika diantaranya yaitu *node rule engine* dan setelah itu dipasangkan ke *node number to string* sebagaimana di paparkan pada Gambar 2.4 dibawah ini:



Gambar 2.4 Fungsi logika

Alur kerja/*Workflow* selanjutnya dalam *data transformation*, terdapat simpul-simpul/*node* yang dipasangkan sebagai mengubah tipe data menggunakan fungsi logika pada gambar 2.5 dibawah ini:



Gambar 2.5 Data Transformation

2.2 Penelitian Terkait

Penelitian sebelumnya telah dilakukan menggunakan *machine learning* dalam topik prediksi cuaca berdasarkan data *time series* informasi cuaca atau iklim dalam beberapa wilayah di dunia. Implementasi *machine learning Decision Tree*[8] , *KNN*[3], *ANN*[7], *SVM*[5], dan analisa regresi[4] telah berhasil dilakukan untuk memprediksi cuaca.

Berikut ini adalah tabel penelitian dari berbagai macam sumber yang penulis gunakan seperti pada Tabel 2.1:

Tabel 2.1 Tabel Penelitian Terkait

No	Judul Penelitian	Metode	Alat Bantu	Hasil
1.	Penerapan Data Mining dengan Metode <i>Decision Tree</i> untuk Prediksi Cuaca di Kota Seattle menggunakan Aplikasi Weka Syarla Shalsabilla Putri Rachmawati1, [8] Tahun 2023	<i>Decision Tree</i>	Weka	85,69%

2.	Penerapan Algoritma K-Nearest Neighbor (KNN) Dalam Memprediksi Dan Menghitung Tingkat Akurasi Data Cuaca Di Indonesia Muhammad Yusuf Rizqon Rangkuti1, Muhamamrd Valensyah Alfansyuri2, Wawan Gunawan3 Tahun 2021	<i>K-Nearest Neighbor</i> (KNN)	Python (Google Colab)	89% [3]
3.	Implementasi <i>Machine learning</i> Dalam Prediksi Cuaca Maritim Dengan Menggunakan Algoritma Artificial <i>Neural Network</i> Lavenberg- Marquardt Kartika Tri Wahyu Ramadhan[7] Tahun 2019	<i>Neural Network</i> Lavenberg- Marquardt	Lavenberq- Marquadt	MSE = 0.026385 [7]
4.	Pendekatan Data Mining Untuk Analisa Curah Hujan Menggunakan Metode Regresi Linear Berganda (Studi Kasus: Kabupaten Aceh Utara) Maya Maulital1*, Nurdin2 [4] Tahun 2023	REGRESI LINEAR	SPSS	nilai koefisien determinasi = 11.8%
5.	Implementasi <i>Machine learning</i> Untuk Memprediksi Cuaca Menggunakan <i>Support Vector Machine</i> Moch. Fauzi Raja Mahendra, Nuril Lutvi Azizah dan Sumarno[5] Tahun 2024	<i>Support Vector Machine</i>	Python	83%

Penelitian terkait pertama yang ditulis oleh (Syarla Shalsabilla Putri Rachmawati1) membahas tentang Algoritma *Decision Tree*. Dari penelitian ini menghasilkan prediksi cuaca yang dirancang berhasil dengan baik, menggunakan tingkat keakuratan kurang lebih 85,69%. dari 1461 data cuaca yang dipelajari, 1252

data berhasil diprediksi menggunakan sahih, serta 209 data diprediksi menggunakan kesalahan. Ini memberikan bahwa modelnya bekerja menggunakan baik, tapi masih ada beberapa masalah yang tidak sempurna. Metode pengujian yang digunakan adalah akurasi dan *Confusion Matrix*[8].

Kemudian penelitian terkait yang kedua ditulis oleh (Muhammad Yusuf Rizqon Rangkuti¹, Muhamamd Valensyah Alfansyuri², Wawan Gunawan³) membahas tentang Algoritma *K-Nearest Neighbor* (KNN). Asal penelitian ini membuat suatu prediksi dengan taraf akurasi data sebanyak 0,8993 atau lebih kurang 89%. menggunakan taraf akurasi data sebesar 89% diharapkan bisa membantu memprediksi suhu dan cuaca pada Indonesia, sebagai akibatnya dapat membantu peternak serta petani buat mengurangi resiko dari gagal panen serta kerugian. Metode pengujian yang digunakan adalah menggunakan akurasi.

Penelitian terkait ketiga ditulis oleh (Kartika Tri Wahyu Ramadhani) membahas tentang Algoritma Artificial Neural Network. Dari penelitian ini menghasilkan peformansi prediksi kecepatan angin sesuai variasi target error sebesar 0.026385 menggunakan Lavenberg- Marquardt. Metode pengujian yang digunakan adalah *Mean Squared Error* (MSE).

Penelitian terkait keempat ditulis oleh (Maya Maulita, Nurdin) membahas tentang algoritma Regresi Linear Berganda. Dari penelitian ini menghasilkan nilai RMSE di variabel suhu (X1) yaitu 77,42mm serta di variabel kelembapan udara (X2) sebesar 77,13mm. Prediksi curah hujan dengan variabel suhu menggunakan metode regresi linear sederhana menghasilkan nilai yang lebih baik daripada perhitungan nilai hubungan memakai variabel suhu dan kelembapan udara yang memakai metode regresi linear berganda. Metode pengujian yang dipergunakan merupakan *root mean squared error* (RMSE).

Penelitian terkait kelima ditulis oleh (Moch. Fauzi Raja Mahendra, Nuril Lutvi Azizah dan Sumarno) membahas tentang algoritma *Support Vector Machine*. Dari penelitian ini menghasilkan pembagian terstruktur mengenai prediksi hujan dengan skor akurasi sebanyak 83% serta precision yang di- hasilkan 0.79% sehingga bila di lihat berasal gap antara skor akurasi dan *precision* contoh tidak mengalami *overfitting* juga *underfitting*. Dengan kata lain penelitian ini menghasilkan model

yang fit dan mendapatkan hasil pengujian sinkron menggunakan realisasi yang pada harapkan. Kelebihan dari metode SVM ini mampu menghasilkan contoh pembagian terstruktur mengenai yang sangat baik. Metode pengujian yang digunakan adalah *Confusion Matrix*.

2.3 Posisi Penelitian

Penelitian ini melakukan analisis terhadap cuaca di Indonesia khususnya di wilayah Jakarta, Bogor, dan Tangerang dengan menggunakan *platform* KNIME. *Platform* KNIME adalah perangkat lunak ilmu data *Open Source* yang memfasilitasi implementasi *machine learning* dan alat penambangan data, asosiasi data beserta visualisasi data. Penelitian yang dilakukan berbeda dengan penelitian sebelumnya karena menggunakan model *machine learning Decision Tree* dengan alat bantu KNIME *Analytics Platform* dalam memprediksi cuaca di wilayah Jakarta, Bogor, dan Tangerang dengan dataset dari tahun 2010-2020. *Confusion Matrix* untuk mengukur dan memahami kinerja model *machine learning*. Selain itu, akurasi juga untuk mengukur seberapa sering model memprediksi kelas dengan benar, baik kelas positif maupun negatif. Semakin tinggi nilai akurasinya, semakin baik model dalam membuat prediksi yang akurat.

Table 2.2 Posisi Penelitian

Penelitian	Prediksi Cuaca	Machine learning	Decision Tree	KNIME
Syarla Shalsabilla Putri Rachmawati1, Tahun 2023	✓	✓	✓	
Muhammad Yusuf Rizqon Rangkuti1, Muhamamad Valensyah Alfansyuri2 , Wawan Gunawan3 Tahun 2021	✓	✓		
Maya Maulita1*, Nurdin2 [4] Tahun 2023	✓	✓	✓	

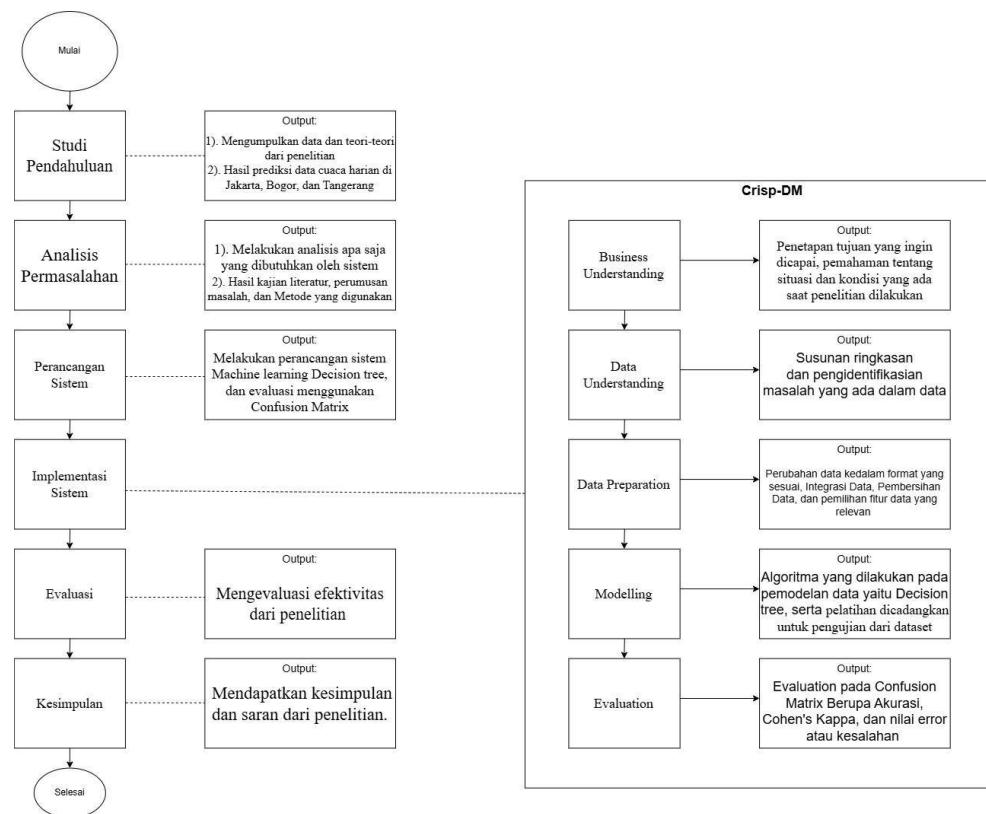
Moch. Fauzi Raja Mahendra, Nuril Lutvi Azizah dan Sumarno Tahun 2024	✓	✓		
H. Celik and A. Cinar 2021		✓	✓	✓
Raafi Alfarizi Tahun 2025	✓	✓	✓	✓

BAB III

ANALISIS DAN PERANCANGAN SISTEM

3.1 Tahapan Penelitian

Berikut ini adalah tahapan-tahapan yang akan dilakukan dalam penulisan ini secara umum:



Gambar 2.6 Tahapan Penelitian

Pada Gambar 2.6 tahapan penelitian dilakukan sebagai berikut:

3.1.1 Studi Pendahuluan

Pada tahap ini, berbagai teori tentang pengembangan *machine learning Decision Tree* cuaca di Jakarta, Bogor, dan Tangerang dikumpulkan dan dianalisis. Metode crisp-

dm digunakan, Tujuannya yaitu untuk menemukan pola yang memikat dan bermakna dalam data yang digunakan. CRISP-DM mempunyai kerangka bertahap dan terstruktur yang memungkinkan pengguna metode lebih fokus dan memahami langkah-langkah yang harus dilakukan dalam suatu penelitian.

3.1.2 Analisis Permasalahan

Berikut ini adalah yang dilakukan penulis pada analisis permasalahan adalah sebagai berikut:

1. Melakukan analisis apa saja yang dibutuhkan oleh sistem.
2. Hasil kajian literatur, perumusan masalah, dan metode yang digunakan.

3.1.3 Perancangan Sistem

Workflow/alur kerja yang dibuat yaitu dengan menghubungkan *node-node* dari ujung ke ujung dalam aplikasi *machine learning* yang dipelajari dalam studi ini. KNIME membaca data "*File Reader*" dan kemudian memprosesnya pada kecepatan yang ditetapkan ke *node "Partitioning"*, di mana 80% pelatihan dicadangkan untuk pengujian 20%. 80% dari set data yang dicadangkan untuk pelatihan diberikan kepada *node Decision Tree Learner*. Pada penghujung algoritma *node Decision Tree Learner* dihubungkan ke *node Predictor* masukan pertama yang memiliki dua masukan dan masukan lain dari *Decision Tree Predictor* mengambil data dari simpul *Partitioning*. Dari data terlatih (*data training*) tersebut, *Decision Tree Predictor* sekarang diperlukan untuk memprediksi data uji.

3.1.4 Implementasi Sistem

Implementasi sistem yang dilakukan oleh penulis yaitu menggunakan Crisp-DM. Adapun tahapan Crisp-DM yang dilakukan oleh penulis untuk hasil Implementasi sistem adalah sebagai berikut:

1. *Business Understanding*

Output dari *business understanding* berupa penetapan tujuan yang ingin dicapai, pemahaman tentang situasi dan kondisi yang ada saat penelitian dilakukan.

2. *Data Understanding*

Berikut ini merupakan *Output* dari *data understanding* yaitu berupa susunan ringkasan dan pengidentifikasi masalah yang ada dalam data.

3. *Data Preparation/Preprocessing Data*

Pada tahap ini, penulis melakukan perubahan data kedalam format yang sesuai, integrasi data, melakukan pembersihan data, dan pemilihan fitur data yang relevan.

4. *Modelling*

Output pada *modelling* berupa algoritma yang dilakukan pada pemodelan data yaitu *Decision Tree*, serta pelatihan yang dicadangkan untuk pengujian dari dataset.

5. *Evaluation*

Dalam mengevaluasi model penelitian ini, menggunakan *Confusion Matrix* Berupa Akurasi, *Cohen's Kappa*, dan nilai error atau kesalahan serta *wrong* dan *correct*.

3.1.5 Evaluasi

Hasil pada pengujian menyatakan bahwa tingkat akurasi yang dimiliki oleh sistem telah mencapai tingkat akurasi yang tinggi dalam memprediksi dan mengklasifikasi cuaca. Nilai pada *Confusion Matrix*, *accuracy*/akurasi, dan *cohen's kappa*, maka jika semakin tinggi nilai tersebut semakin dapat diandalkan model klasifikasi kita untuk membuat prediksi yang akurat pada data baru yang belum pernah dilihat sebelumnya.

3.1.6 Kesimpulan

Adapun kesimpulan dari penelitian ini yaitu sebagai berikut:

1. Dapat menjawab rumusan masalah dari penelitian ini.

2. Mendapatkan saran dari kesimpulan yang dibuat berdasarkan rumusan masalah

3.2 Jenis Penelitian

Penelitian ini menggunakan metode *Research and Development*, karena dalam penelitian melakukan analisis *data mining* dan pemodelan *machine learning* algoritma *Decision Tree* berdasarkan dataset *time series* data cuaca dan iklim di wilayah Jakarta, Bogor, dan Tangerang. Dalam penerapan metode penelitian *research and development* ini menggunakan tahapan-tahapan yang terdapat dalam metode crisp-dm dengan alat pengembangan KNIME *Analytics Platfrom*.

3.3 Metode Analisis Data

Penelitian ini akan melakukan pengujian model dengan pendekatan analisis data secara kuantitatif. Metode ini berfokus pada data *time series* berupa angka/numerik atau data nominal yang dapat diukur. Dalam penelitian ini, hasil analisis dalam bentuk statistik atau angka-angka tertentu yang dihasilkan selanjutnya di visualisasikan menggunakan *platform* KNIME.

3.4 Metode Pengumpulan Data

Penelitian ini dilakukan pengumpulan data melalui studi pustaka dengan mengumpulkan data artikel-artikel penelitian terkait penerapan *machine learning* dan *data mining* dalam topik untuk memprediksi cuaca atau iklim dunia. Selain itu pengumpulan data juga dilakukan dengan mengambil sampel data yang tersedia di *internet*.

Sumber dataset penelitian ini diperoleh melalui BMKG yang di *share* melalui platform aplikasi online [kaggle.com](https://www.kaggle.com). Dataset yang digunakan berasal dari data *time series* cuaca dan iklim di Indonesia dari tahun 2010-2020. Pada data ini digunakan untuk mengetahui data yang akan diprediksi. Data lapangan yang memberikan informasi terkait pada data cuaca harian di Jakarta, Bogor, dan Tangerang.

3.5 Metode Pengujian

Pada penelitian ini, Berdasarkan umpan balik pengguna atau user, yang dilakukan untuk pengujian dengan menggunakan *node Scorer* yang dimana terdapat *Confusion Matrix*. *Confusion Matrix* adalah alat yang digunakan untuk mengevaluasi kinerja model klasifikasi. Ini merupakan tabel perbandingan prediksi model dengan nilai sebenarnya (kondisi nyata). Dengan menggunakan *confusion matrix*, kita dapat menghitung berbagai metrik penting seperti akurasi, presisi, perolehan, dan skor F1, yang dapat memberi kita gambaran tentang seberapa baik model klasifikasi kita bekerja. Selain itu, *Accuracy*/akurasi berfokus pada mengukur seberapa sering model memprediksi kelas dengan benar, baik kelas positif maupun negatif. Semakin tinggi nilai akurasinya, semakin baik model dalam membuat prediksi yang akurat. Namun, akurasi saja tidak selalu cukup, terutama jika dataset memiliki ketidakseimbangan kelas (*class imbalance*). Dalam kasus seperti itu, akurasi bisa menjadi metrik yang menyesatkan. Oleh karena itu, perlu mempertimbangkan metrik evaluasi lain seperti presisi, *recall*, dan skor F1 secara bersamaan dengan akurasi. Sementara itu, *Cohen's kappa* digunakan untuk menunjukkan seberapa akurat prediksi model kita dibandingkan dengan hasil sebenarnya. Berikut ini adalah tabel untuk *Confusion Matrix* yang dijelaskan pada Tabel 3.5.1:

Tabel 3.5.1 Tabel *Confusion Matrix*

		Prediksi	
Aktual	True	False	
True	TP	FP	
False	TN	FN	

Lalu berikut ini adalah penjelasan berupa tabel pada *cohen's kappa* yang dijelaskan pada Tabel 3.5.2 dibawah ini:

Tabel 3.5 2 Tabel *Cohen's Kappa*

Statistik <i>Cohen's Kappa</i>	Tingkat Kesepakatan
< 0.00	Buruk
0.00 – 0.20	Sedikit Buruk
0.21 – 0.40	Cukup
0.41 – 0.60	Sedang
0.61 – 0.80	Kuat
0.81 – 0.99	Hampir Mendekati Sempurna
1.00	Sempurna

3.6 Evaluasi Hasil Pengujian

Hasil pada pengujian menyatakan bahwa tingkat akurasi yang dimiliki oleh sistem telah mencapai tingkat akurasi yang tinggi dalam memprediksi dan mengklasifikasi cuaca. Nilai pada *Confusion Matrix*, *accuracy*/akurasi, dan *cohen's kappa*, maka jika semakin tinggi nilai tersebut semakin dapat diandalkan model klasifikasi kita untuk membuat prediksi yang akurat pada data baru yang belum pernah dilihat sebelumnya.

3.7 Lingkungan Pengembangan

a. Tempat

Penelitian ini dilakukan pada BMKG cuaca yang berlokasi di wilayah Jakarta, Bogor, dan Tangerang.

b. Alat dan bahan

Adapun alat yang digunakan penulis untuk penelitian ini adalah:

- Spesifikasi *hardware* yang digunakan yaitu laptop merek *Lenovo* yang memiliki

RAM 8GB

Dan *software* yang digunakan penulis pada penelitian ini adalah:

- *KNIME Analytics Platform*
- *Microsoft Excel*
- *Draw.io*
- *Mendeley Dekstop*
- *Google Chroome*

BAB IV

IMPLEMENTASI DAN EVALUASI

Pada bagian ini akan dijelaskan mengenai implementasi algoritma *decision tree* berdasarkan Crisp-DM yang meliputi *business understanding*, *data understanding*, *data preparation*, *modelling*, dan *evaluation*.

4.1 Business Understanding

Penerapan *data mining* pada penelitian ini berhubungan eksklusif menggunakan data suhu rata-rata untuk menggali pengetahuan pembahasan suatu pola terhadap intensitas suhu minimum dan maksimum yang berpotensi terhadap curah hujan, kelembaban, kecepatan angin, dan durasi penyinaran matahari. Tujuan utamanya adalah buat mengeksplorasi bagaimana suhu rata-rata suatu wilayah terbentuk, serta mengidentifikasi faktor-faktor yang mempunyai pengaruh signifikan terhadap naik turunnya suhu tersebut. Fokus spesifik diberikan di dua komponen krusial pembentuk suhu rata-rata, yaitu suhu minimum dan maksimum. Kedua parameter tersebut dipertimbangkan karena secara teoritis dan mudah sangat mungkin memengaruhi fluktuasi suhu harian serta bulanan.

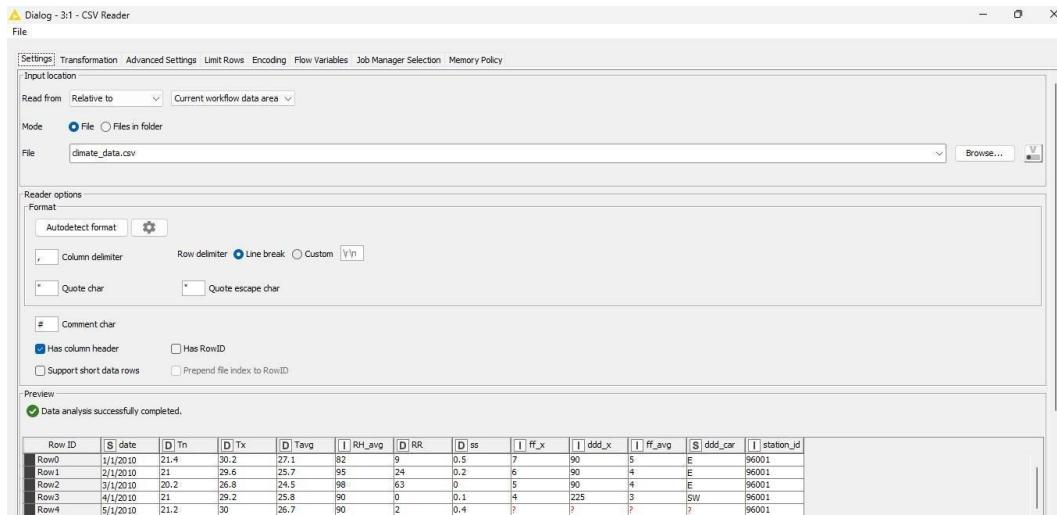
4.2 Data Understanding

Penelitian ini menggunakan data yang digabungkan antara *climate_data.csv* dan *station_detail.csv* yang bersumber dari Badan Meteorologi Klimatologi dan Geofisika (BMKG). Data tersebut merupakan data cuaca periode tahun 2010 sampai dengan tahun 2020. Jumlah seluruh dataset awal 589.265 mencakup seluruh wilayah Indonesia, dan untuk wilayah Jakarta, Bogor, dan Tangerang jumlah datanya adalah 4.524. Pada Tabel 4.1 adalah ditampilkan jenis kolom dari kedua sumber data.

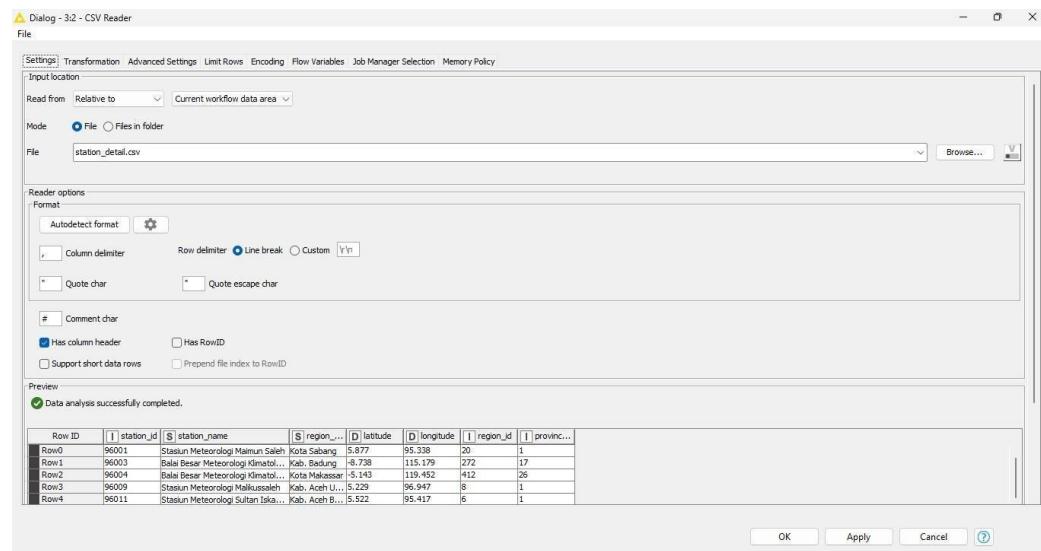
Tabel 4. 1 Jenis kolom dari sumber data

<i>Node 1 (climate_data.csv)</i>	<i>Node 2 (station_detail.csv)</i>
<ol style="list-style-type: none"> 1. Date (Tanggal) 2. Tn (Suhu Minimum) 3. Tx (Suhu Maksimum) 4. Tavg (Suhu Rata-rata) 5. RH_avg (Kelembaban Rata-rata) 6. RR (Curah Hujan) 7. Ss (Durasi Penyinaran Matahari) 8. Ff_x (Kecepatan Angin Maksimum) 9. Ddd_x (Arah Angin pada Kecepatan Maksimum) 10. Ff_avg (Kecepatan Angin Rata-rata) 11. Ddd_car (Arah Angin Tertinggi) 	<ol style="list-style-type: none"> 1. Station_id 2. Station_name 3. Region_name 4. Latitude 5. Longitude 6. Region_id 7. Province_id

Indikator data cuaca yang digunakan yaitu suhu minimum, suhu maksimum, curah hujan, durasi penyinaran matahari, kelembababan rata-rata, kecepatan angin maksimum dan kecepatan angin rata-rata berdasarkan wilayah dengan klasifikasi suhu rata-rata panas adalah lebih dari 25°C dan dingin kurang dari 25°C[16]. Pada Gambar 4. 1 ditampilkan dataset *climate_data.csv* yang data tersebut berisi cuaca harian disertai dengan nomor stasiun (*station_id*). Dan pada Gambar 4. 2 *station_detail.csv* yang data tersebut berisi detail dari nama-nama stasiun BMKG dilengkapi dengan nomor dan nama stasiun, nama provinsi, nomor wilayah kabupaten atau kota, letak geografis (*latitude* dan *longitude*) dan nomor provinsi.



Gambar 4. 1 Tampilan *node csv reader* untuk membaca data *climate_data.csv*



Gambar 4. 2 Tampilan konfigurasi node 2 untuk membaca data *station_detail.csv*

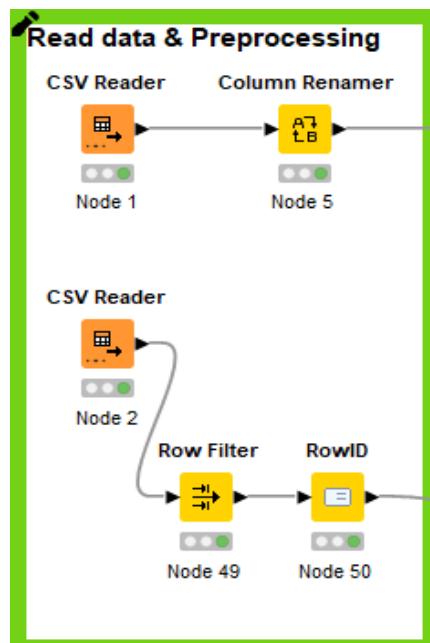
4.3 Data Preparation

Data yang dihasilkan merupakan gabungan dari dua dataset yaitu data cuaca sebanyak 589265 data yakni data cuaca dari Januari 2010 hingga Desember 2020 serta data detail stasiun sebesar 192 data. Terdapat beberapa data yang tidak lengkap atau

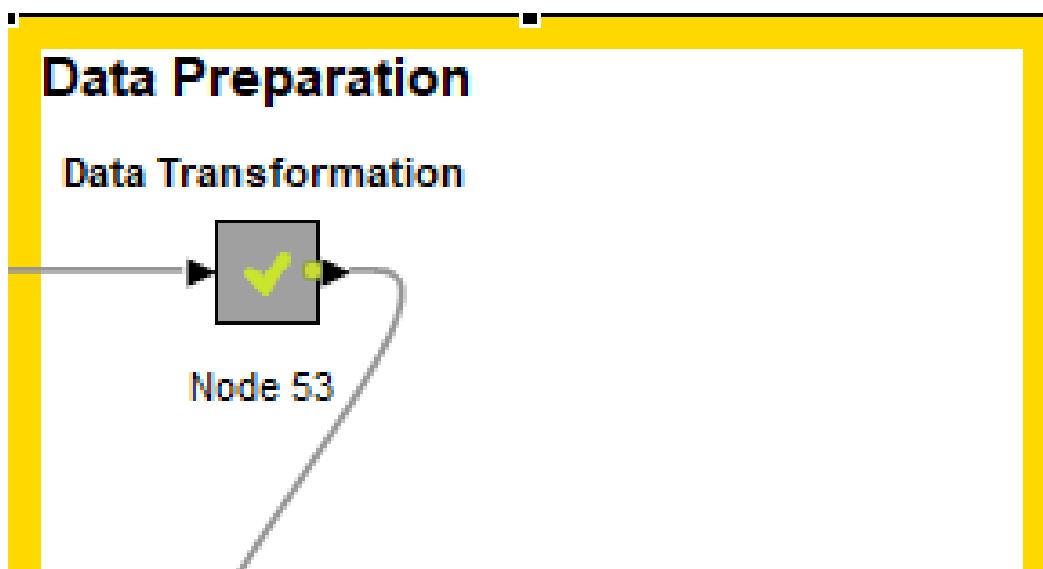
hilang menyebabkan ketidakbenaran serta kekeliruan. Oleh karena itu dilakukan *data transformation*, *data integration*, *data cleaning*, serta *feature selection*[25]. Berikut ini adalah tahapan yang dilakukan penulis untuk *data preparation* sebagai berikut:

4.3.1 Data Transformation

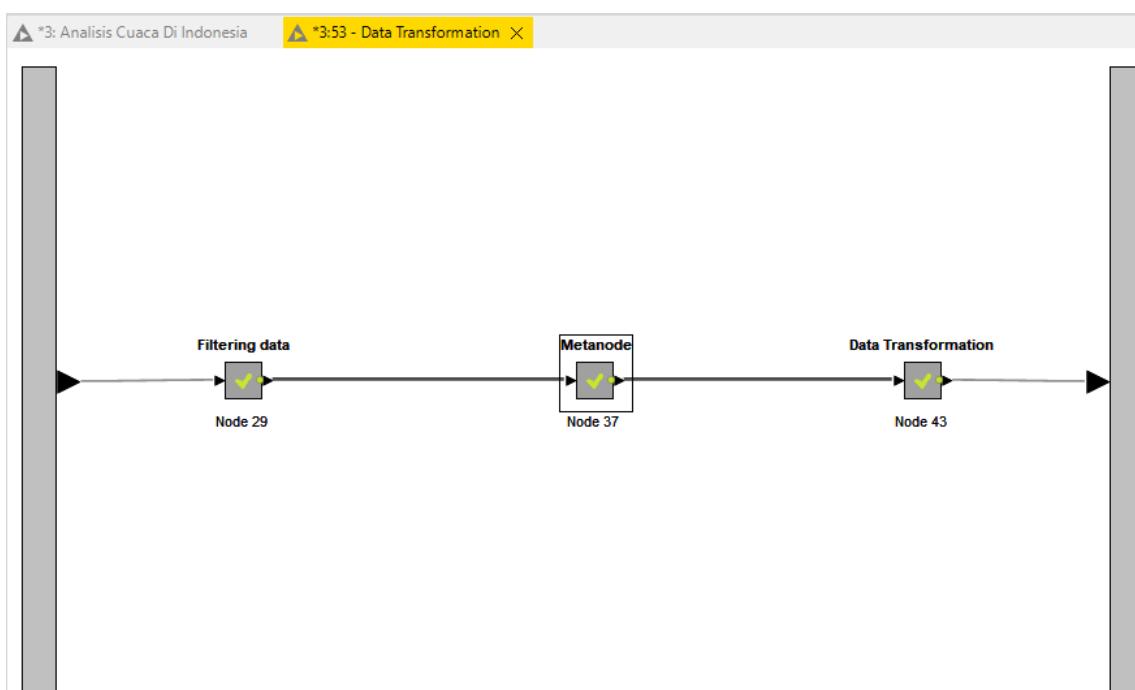
Pada tahap data transformation, yang mana dilakukan perubahan pada atribut label agar sinkron menggunakan kebutuhan pada proses evaluasi dan mengubah nama pada data cuaca sebagaimana sudah dijelaskan di Tabel 4. 1 menggunakan *platform* Knime. Label pada kolom suhu rata-rata yang lebih dari 25°C diubah menjadi ">25", sementara yang kurang dari 25°C diubah menjadi "< 25". Kemudian pada kolom *station_id* yang mana tipe data tersebut adalah *integer* yang diubah menjadi *string* untuk memudahkan penilaian pada contoh yang akan dibuat menggunakan tahapan preprocessing dan data transformation yang telah dilakukan, dataset siap buat diproses memakai prosedur pemecahan *decision tree* yang akan membuat pohon keputusan buat memprediksi cuaca di daerah Jakarta, Bogor, dan Tangerang sesuai atribut-atribut relevan yang telah dipilih sebelumnya[26].



Gambar 4. 3 Data Transformation 1



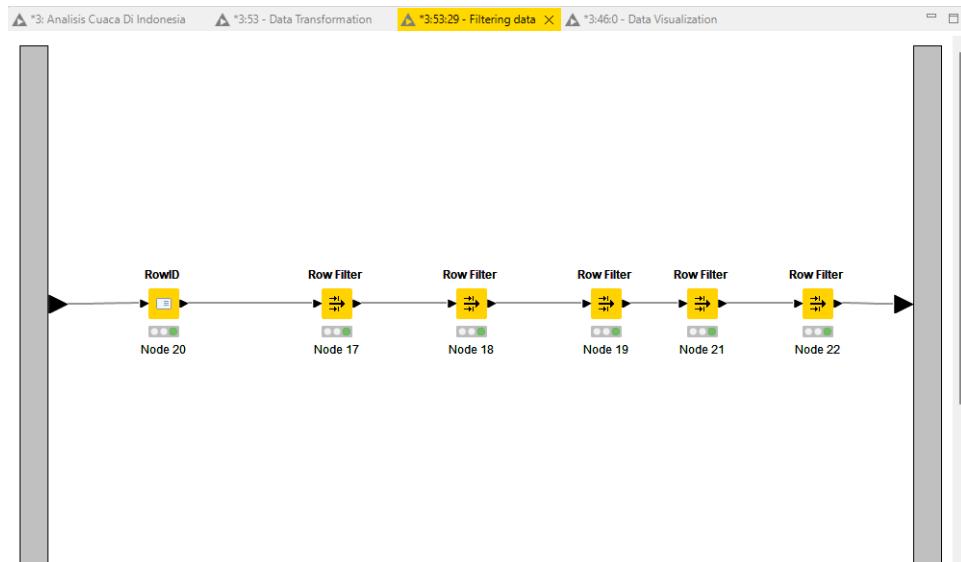
Gambar 4. 4 *Data Transformation 2*



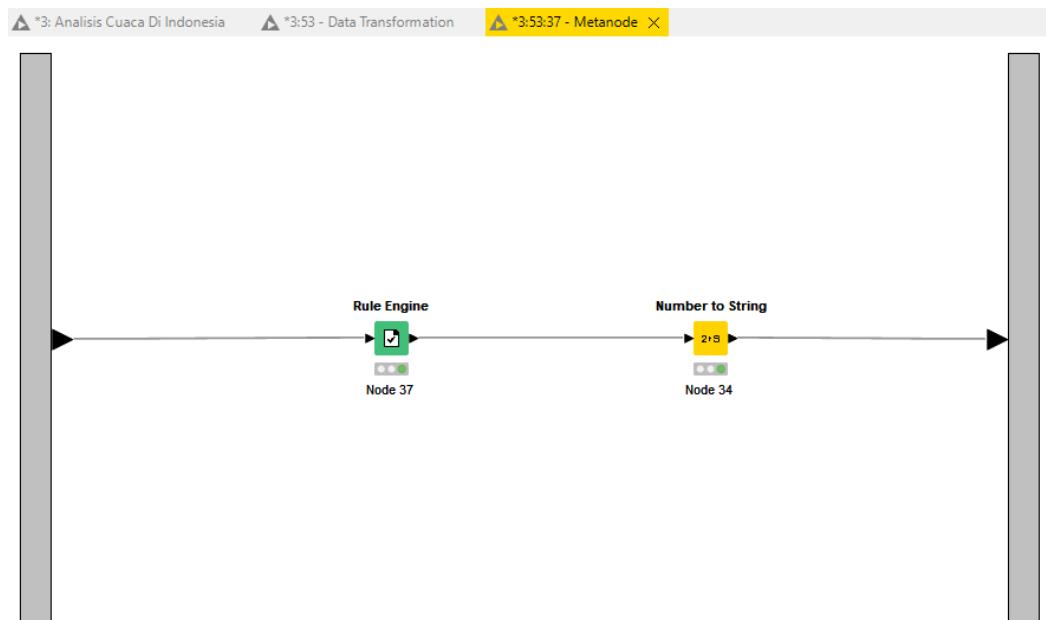
Gambar 4. 5 Alur kerja *data transformation 2*

Penjelasan pada Gambar 4. 3 sampai 4. 5 terdapat beberapa node yang saling berpasangan yaitu 2 buah *node csv reader*, *column renamer*, *row filter* dan *rowid* pada Gambar 4. 3. Pada Gambar 4. 4 merupakan kumpulan *node* (*metanode*) yang berisikan beberapa *metanode* untuk transformasi data sebagaimana yang dipaparkan pada Gambar 4. 5. Selanjutnya pada Gambar 4. 5 terdapat 3 buah *metanode* yang masing-masing berisi

beberapa node yang berpasangan sebagaimana yang akan dipaparkan pada Gambar 4. 6 sampai dengan Gambar 4. 8 dibawah ini:



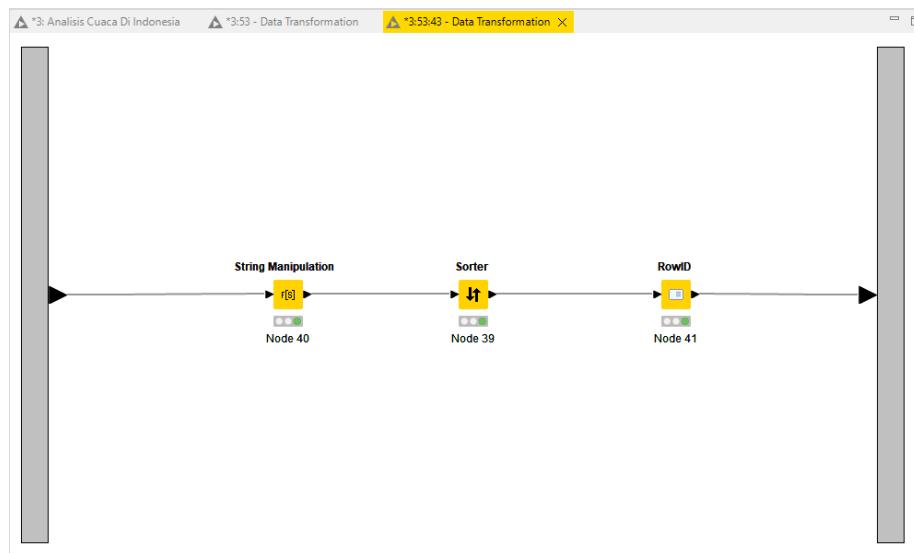
Gambar 4. 6 Alur kerja untuk *filtering data*



Gambar 4. 7 Alur kerja fungsi logika dan perubahan tipe data untuk *data transformation*

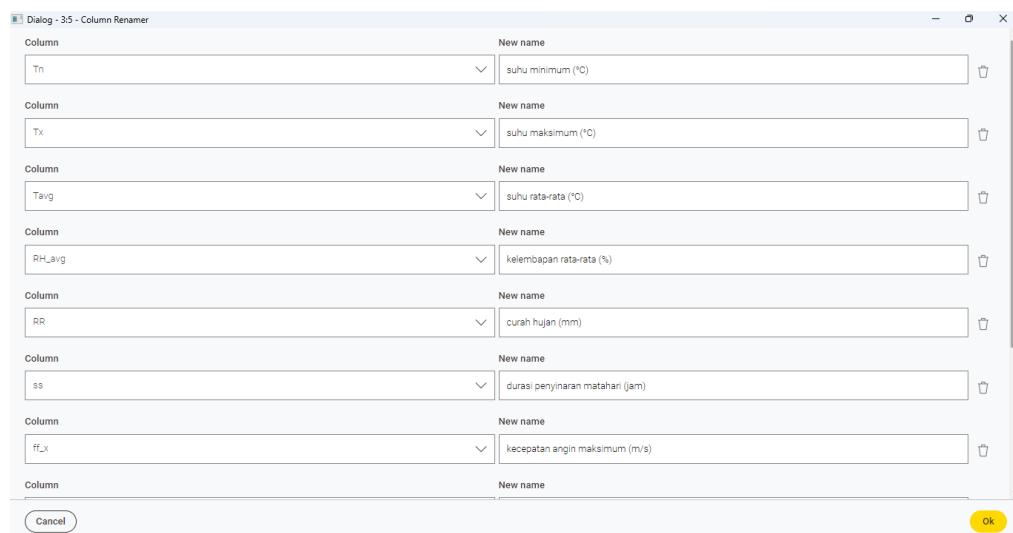
Penjelasan Gambar 4. 7 terdapat 2 buah *node* yang saling berpasangan untuk *data transformation* yaitu *rule engine* dan *number to string*. *Node rule engine* merupakan

fungsi logika yang mana pada penelitian ini digunakan untuk merubah tipe data pada kolom yang dijadikan sebagai target sebagaimana yang dipaparkan pada Gambar 4. 30. Sedangkan *node number to string* digunakan untuk merubah tipe data *number (double)* yang sebenarnya bukan data *numerical* menjadi string.

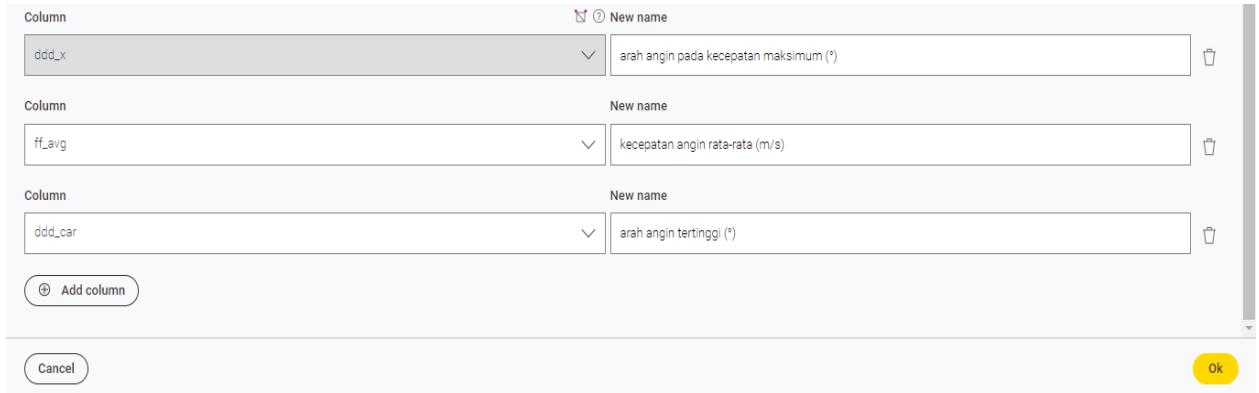


Gambar 4. 8 *Data transformation 3*

Berikut ini adalah langkah-langkah konfigurasi untuk *data transformation* pada tahap *data preparation* yang dipaparkan pada Gambar 4.9 sampai dengan 4. 40:



Gambar 4. 9 Konfigurasi *node column renamer* untuk merubah nama kolom



Gambar 4. 10 Kolom selanjutnya yang akan di *rename*

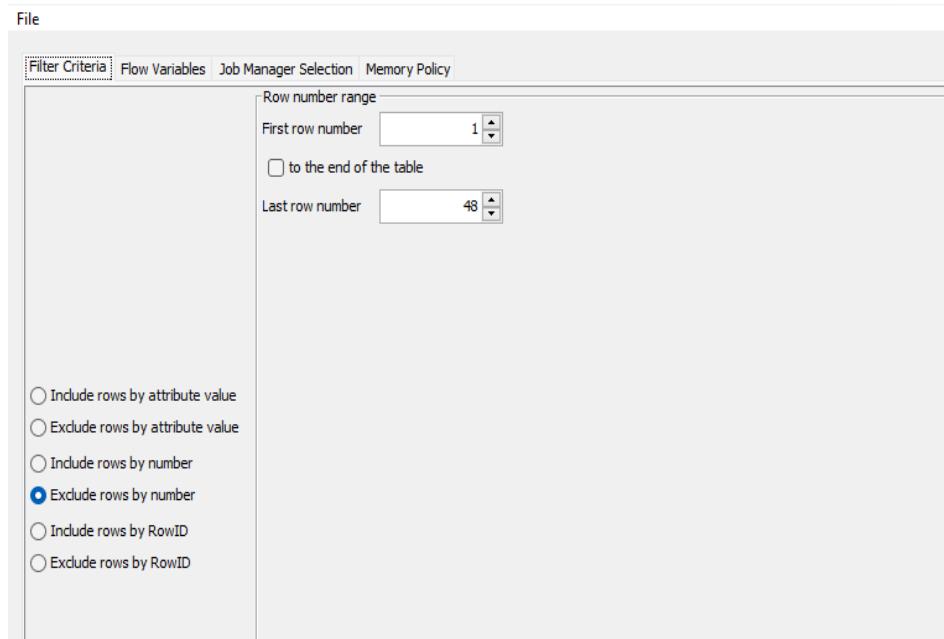
Penjelasan pada Gambar 4. 9 dan Gambar 4. 10 merupakan konfigurasi untuk mengganti nama kolom (*rename*). *Node column renamer* berfungsi untuk merubah atau mengganti nama kolom sesuai dengan apa nama kolom yang kita inginkan. Tujuannya untuk memudahkan pemahaman penulis bahwa kolom mana yang akan dijadikan acuan untuk prediksi melalui pemilihan fitur yang disebut dengan *feature selection*.

Row ID	S date	D suhu minimum (°C)	D suhu maksimum (°C)	D suhu rata-rata (°C)	I kelerapan rata-rata (%)	D curah hujan (mm)	D durasi penyinaran matahari (jam)	I kecepat...	I arah an...	I kecepa...	S arah an...	I
Row0	1/1/2010	21.4	30.2	27.1	82	9	0.5	7	90	5	E	960
Row1	2/1/2010	21	29.6	25.7	95	24	0.2	6	90	4	E	960
Row2	3/1/2010	20.2	26.8	24.5	98	63	0	5	90	4	E	960
Row3	4/1/2010	21	29.2	25.8	90	0	0.1	4	225	3	SW	960
Row4	5/1/2010	21.2	30	26.7	90	2	0.4	?	?	?	?	960
Row5	6/1/2010	21.2	30	26.1	93	11	0.3	?	?	?	?	960
Row6	7/1/2010	21.4	29	25.4	96	2	0.1	?	?	?	?	960
Row7	8/1/2010	21.8	29.8	26.8	91	3	0.6	5	90	4	E	960
Row8	9/1/2010	21.8	30.2	26.7	92	3	0.7	4	90	3	E	960
Row9	10/1/2010	20	30.2	27.1	88	28	0.6	5	90	4	E	960
Row10	11/1/2010	20	30.2	25.7	92	4	0.6	4	90	3	E	960
Row11	12/1/2010	23.4	30.2	27.3	85	0	0.6	3	45	3	NE	960
Row12	13/01/2010	21.8	30.2	26.9	82	0	0.2	4	90	3	E	960
Row13	14/01/2010	22	30	26.3	87	3	0.6	11	90	4	E	960
Row14	15/01/2010	22.4	29.8	26.4	88	1	0.4	8	90	7	E	960
Row15	16/01/2010	22.8	29.8	27.4	82	0	0.6	8	90	7	E	960
Row16	17/01/2010	22.2	30.8	27.5	80	0	0.6	8	90	6	E	960
Row17	18/01/2010	21.8	30.4	26.7	89	0	0.6	4	90	3	E	960
Row18	19/01/2010	20.4	29.8	25.9	95	22	0.2	3	90	3	E	960
Row19	20/01/2010	20.4	29.8	26.1	88	0	0.6	3	90	3	E	960
Row20	21/01/2010	20.2	30	25.5	92	0	0.4	?	?	?	?	960
Row21	22/01/2010	20.2	29.8	25.8	92	14	0.6	5	90	4	E	960
Row22	23/01/2010	20.1	30.8	27.2	90	0	0.6	6	90	4	E	960
Row23	24/01/2010	20.2	29	25.9	92	7	0.4	6	90	5	E	960
Row24	25/01/2010	20.2	29.8	26.8	86	0	0.3	8	90	5	E	960
Row25	26/01/2010	22.2	30.4	27.5	86	0	0.6	7	90	5	E	960
Row26	27/01/2010	20	30	26.9	88	3	0.3	9	90	7	E	960
Row27	28/01/2010	21.6	30	26.9	89	6	0.4	7	90	5	E	960
Row28	29/01/2010	21.6	30.2	27.5	85	0	0.6	6	90	5	E	960
Row29	30/01/2010	21	30.4	26.4	89	0	0.6	6	90	4	E	960
Row30	31/01/2010	22	31	26.8	87	0	0.6	5	90	4	E	960
Row31	1/2/2010	23	30	27.2	86	0	0.6	5	90	4	E	960
Row32	2/2/2010	20.4	31	27.5	87	0	0.6	8	90	5	E	960
Row33	3/2/2010	21	30.4	27.2	85	0	0.6	4	90	4	E	960
Row34	4/2/2010	21	30.6	26.6	88	0	0.6	?	?	?	?	960
Row35	5/2/2010	20.8	31	26.5	88	0	0.6	4	90	3	E	960
Row36	6/2/2010	21	31.2	27	84	0	0.6	4	90	3	E	960
Row37	7/2/2010	23.2	31.2	27.5	85	0	0.6	6	90	4	E	960

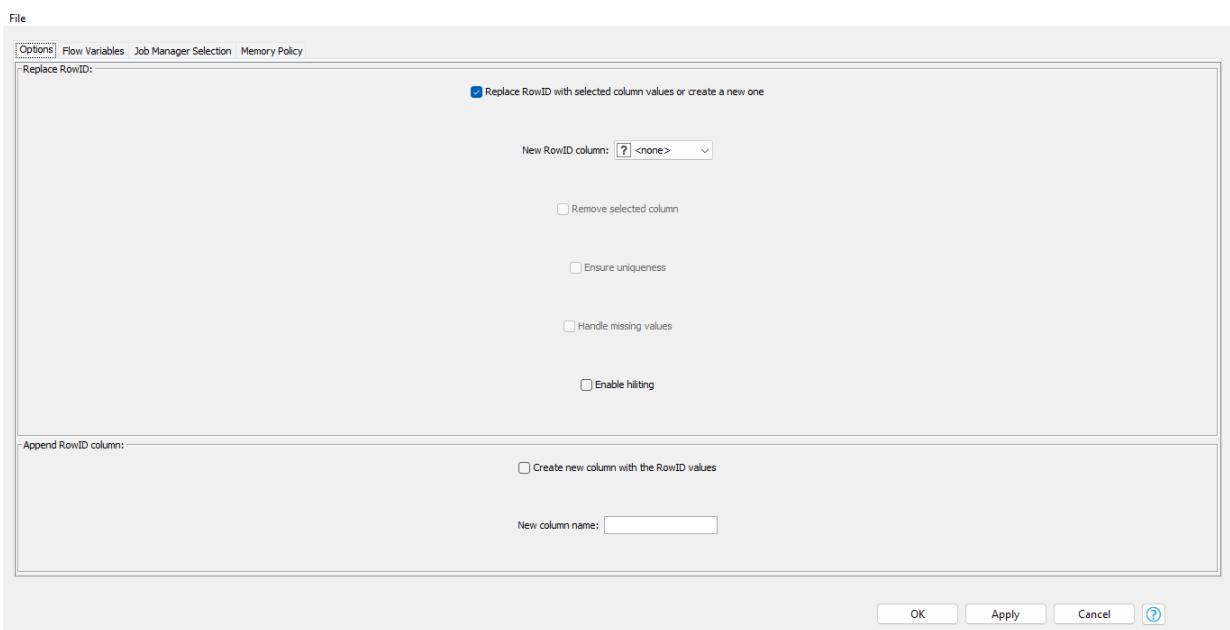
Gambar 4. 11 Tampilan tabel data setelah kolom di *rename*

Penjelasan pada Gambar 4. 10 dan Gambar 4. 11 yaitu konfigurasi *node column renamer* yang digunakan untuk merubah nama kolom (*rename*). Dan pada Gambar 4. 11 yaitu tampilan tabel data pada *node column renamer* sesudah nama kolom diubah

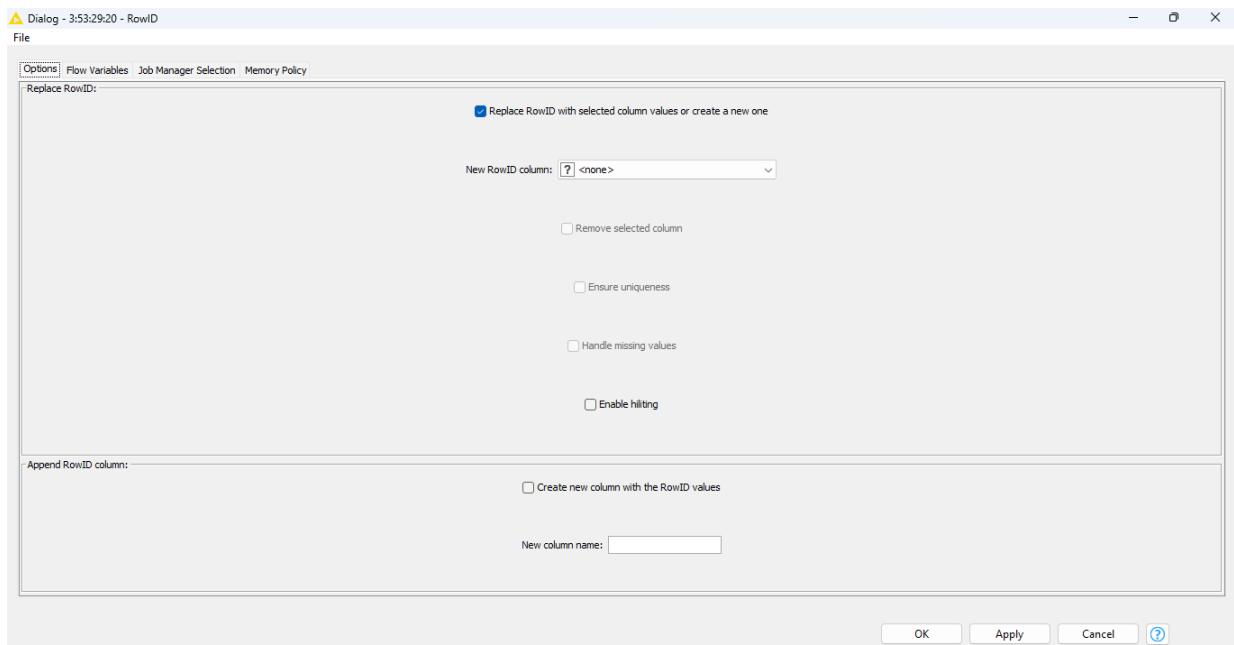
namanya. Kemudian berikut ini pada Gambar 4. 12 adalah baris data *station_detail.csv* yang akan di *exclude* menggunakan *node row filter* dan pada Gambar 4. 13 akan diurutkan kembali data yang telah di *exclude* barisnya:



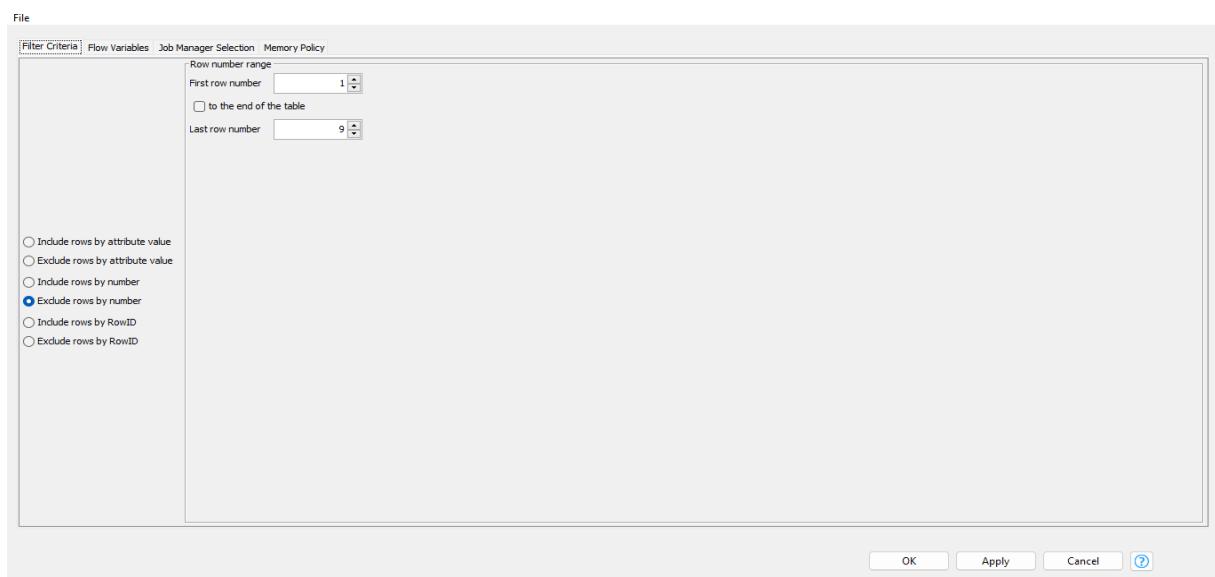
Gambar 4. 12 Konfigurasi *node row filter* untuk *exclude* baris 1-48 yang tidak diperlukan



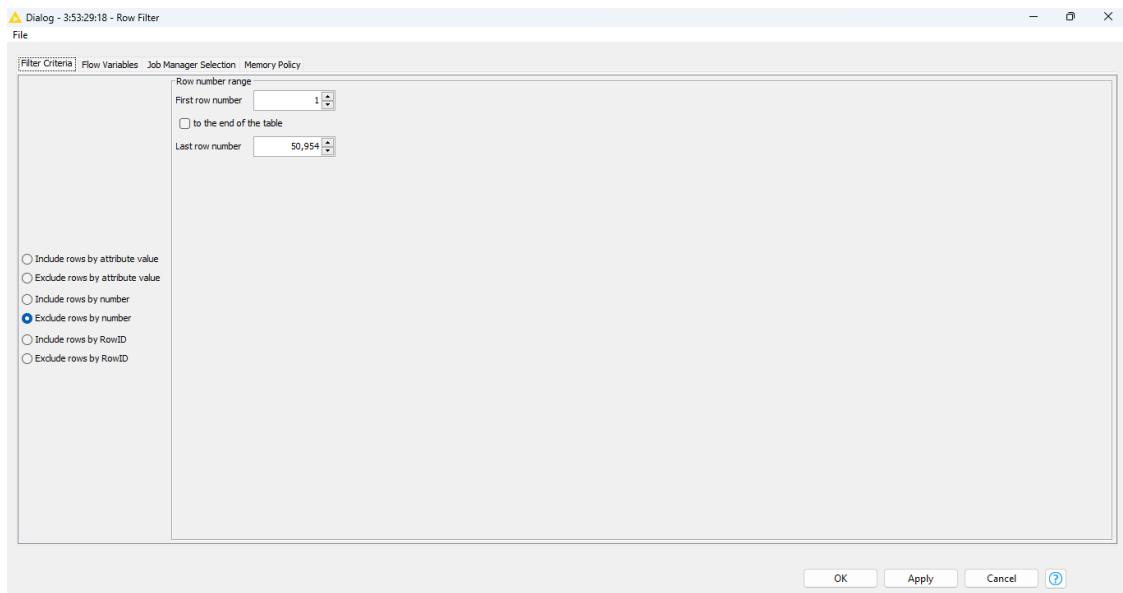
Gambar 4. 13 Konfigurasi *node row id* untuk mengubah kembali baris data *station_detail.csv* yang sebelumnya bukan dari baris 1



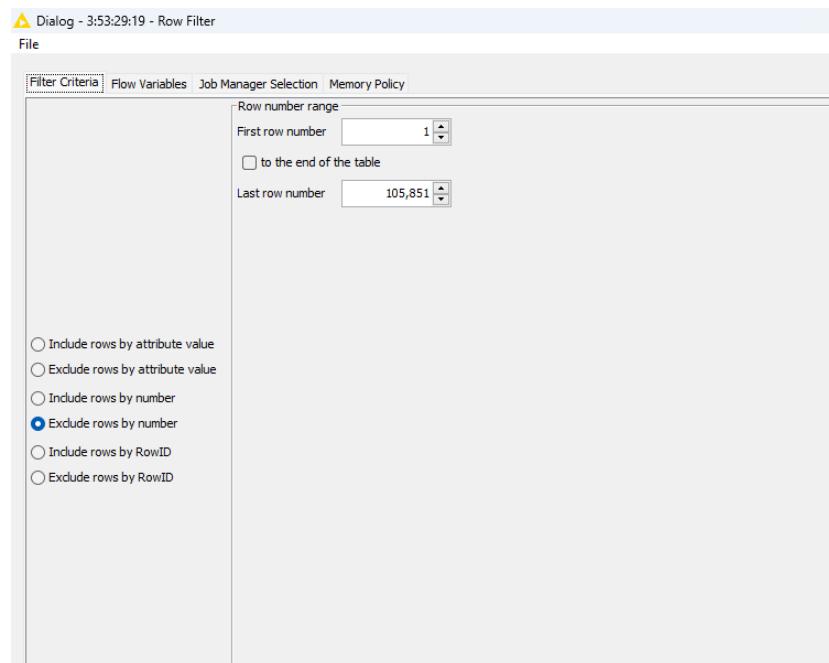
Gambar 4. 14 Konfigurasi untuk mengurutkan data *climate_data.csv*



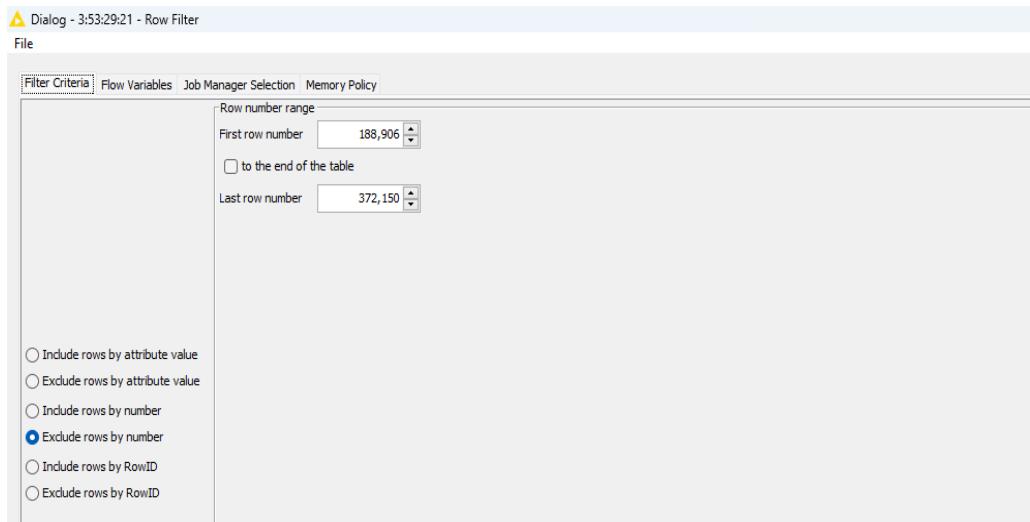
Gambar 4. 15 *Exclude* baris 1-9



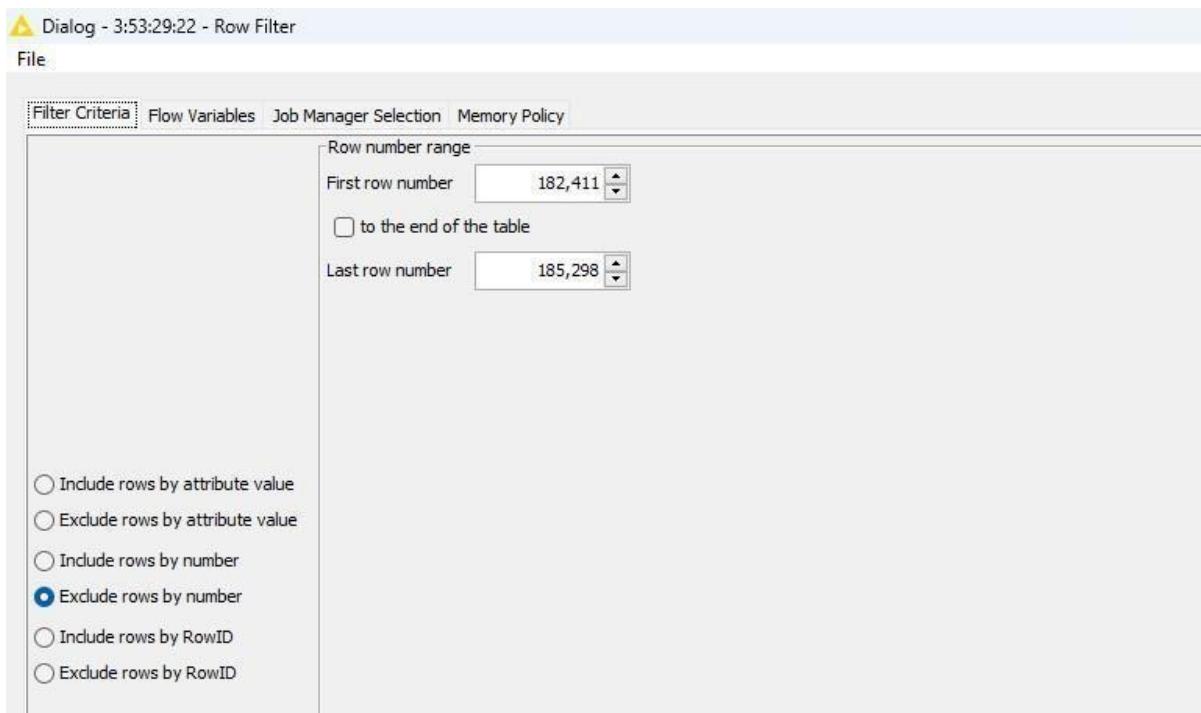
Gambar 4. 16 *Exclude* baris 1 - 50.954



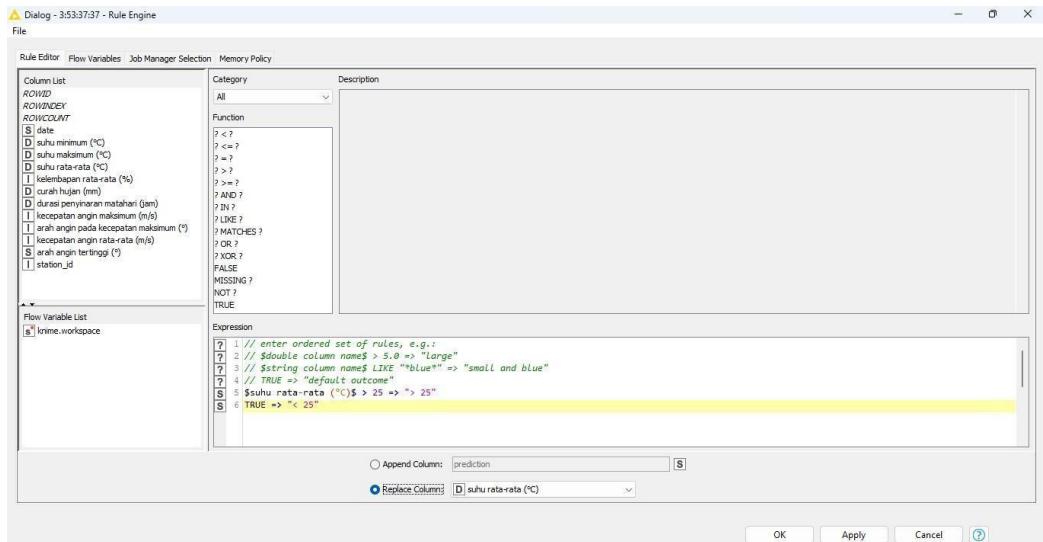
Gambar 4. 17 *Exclude* baris 1 - 105.851



Gambar 4. 18 *Exclude* baris 188.906 - 372.150



Gambar 4. 19 *Exclude* baris 182.411 - 185.298

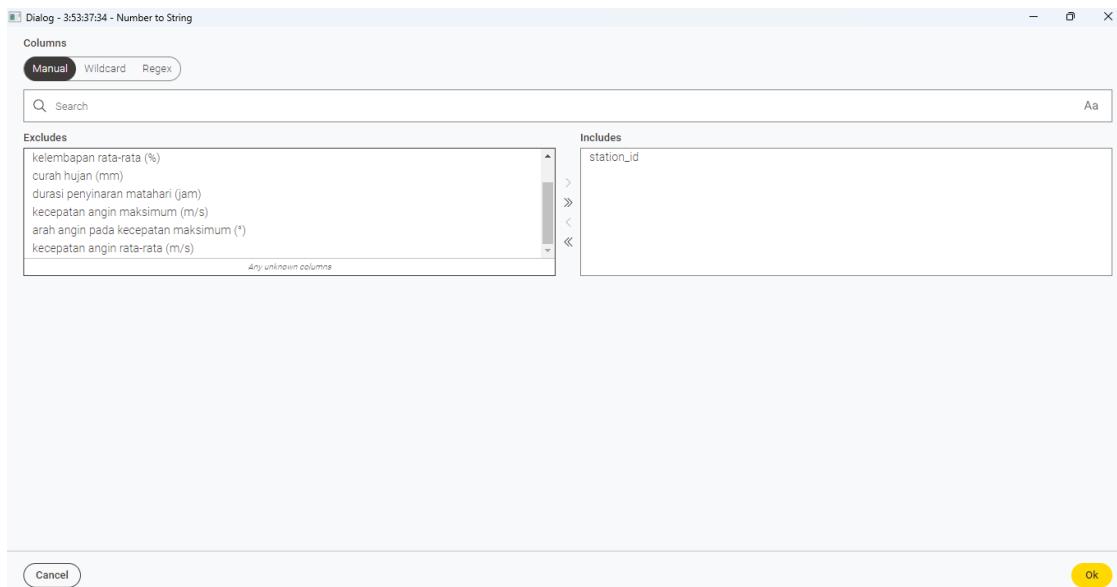


Gambar 4. 20 Kode program/sintax untuk konfigurasi node rule engine

D suhu rata-rata (°C)
27.5
27.9
26.4
26.6
28
27.7
26
25.9
26
26.2
27.2
?
26.6
25.9

Gambar 4. 21 Tampilan isi kolom sebelum di konfigurasi

Gambar 4. 22 Tampilan isi baris pada kolom setelah dikonfigurasi



Gambar 4. 23 Konfigurasi *node number to string*

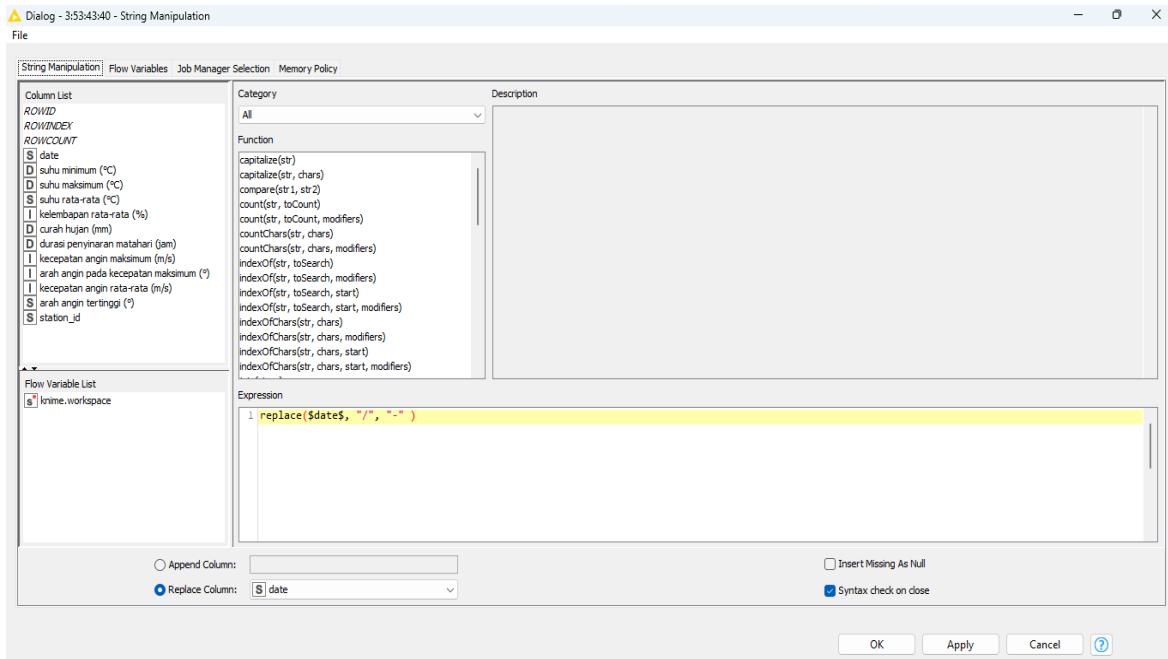
I	station_id
	96297
	96297
	96297
	96297

Gambar 4. 24 Kolom data sebelum di konfigurasi *number to string*

S	station_id
	96297
	96297
	96297
	96297
	96297
	96297
	96297
	96297

Gambar 4. 25 Kolom data sesudah di konfigurasi menggunakan *number to string*

Penjelasan pada gambar 4. 20 sampai dengan Gambar 4. 25 adalah proses *data transformation* yang dimana mengubah tipe data yang sebelumnya *double* menjadi *string*. Tipe data yang akan kita ubah yaitu suhu rata-rata yang bertipe data *double* menjadi *string*, karena kolom data tersebut akan menjadi acuan untuk pengambilan keputusan dari penelitian ini. Selain kolom suhu rata-rata, terdapat beberapa kolom bertipe data *integer* dan *double* yang akan diubah menjadi *string* diantaranya adalah *station_id*. Hal ini disebabkan karena kolom data tersebut memang berupa angka, tetapi bukan berupa jumlah atau ukuran.

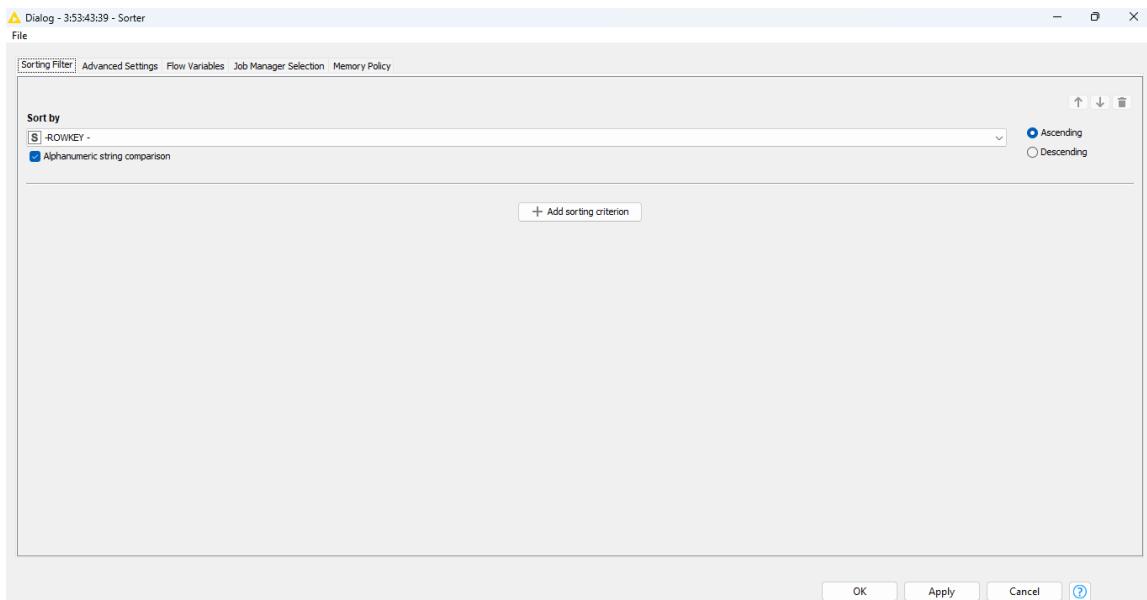


Gambar 4. 26 Konfigurasi *string manipulation*

Penjelasan Gambar 4. 26 merupakan konfigurasi menggunakan *node string manipulation*. *String manipulation* berfungsi untuk memanipulasi atau merubah isi baris menggunakan sintax atau *function*. Kolom data yang akan diubah atau dimanipulasikan barisnya yaitu kolom *date* atau kolom tanggal yang dimana baris pada kolom tersebut campuran antara garis miring (/) dan tanda strip (-). Maka kolom data tersebut akan diubah barisnya menjadi tanda strip (-) saja menggunakan *function* yang dituliskan di *expression*. Sebagaimana yang akan dipaparkan pada Gambar 4. 27 dibawah ini:

S	date
	2-2-2016
	3-2-2016
	4-2-2016
	5-2-2016
	6-2-2016
	7-2-2016
	8-2-2016
	9-2-2016
	10-2-2016

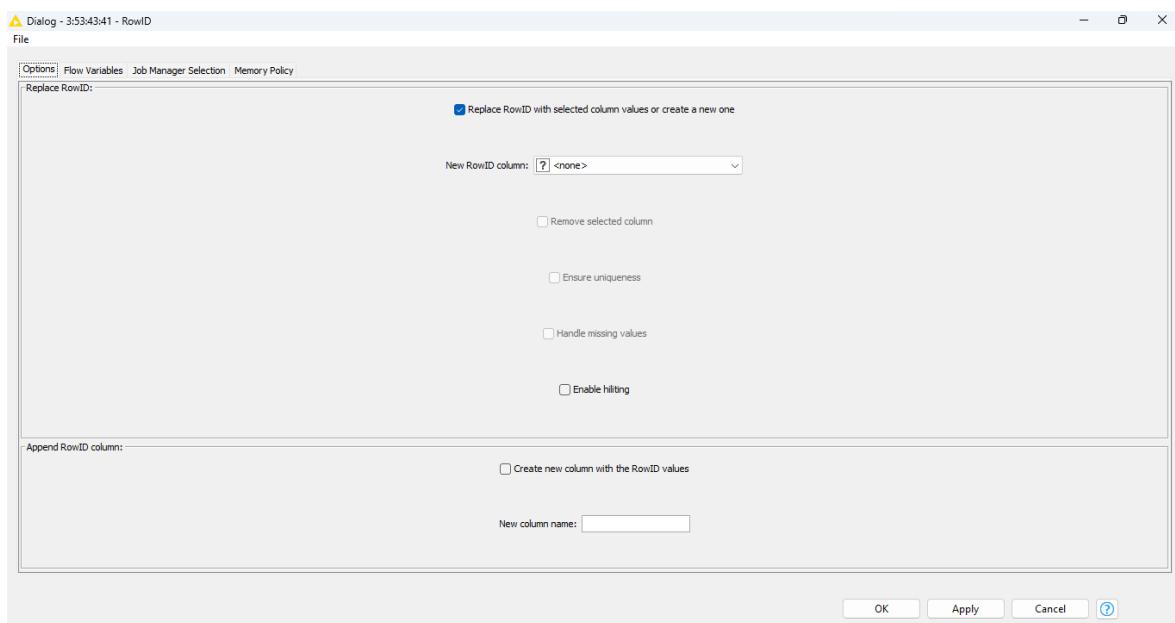
Gambar 4. 27 Tampilan kolom data setelah dikonfigurasi



Gambar 4. 28 Konfigurasi *node sorter* untuk mengurutkan baris dari yang terlama ke terbaru

S	date
	1-1-2010
	2-1-2010
	3-1-2010
	4-1-2010
	5-1-2010
	6-1-2010
	7-1-2010
	8-1-2010
	9-1-2010
	10-1-2010

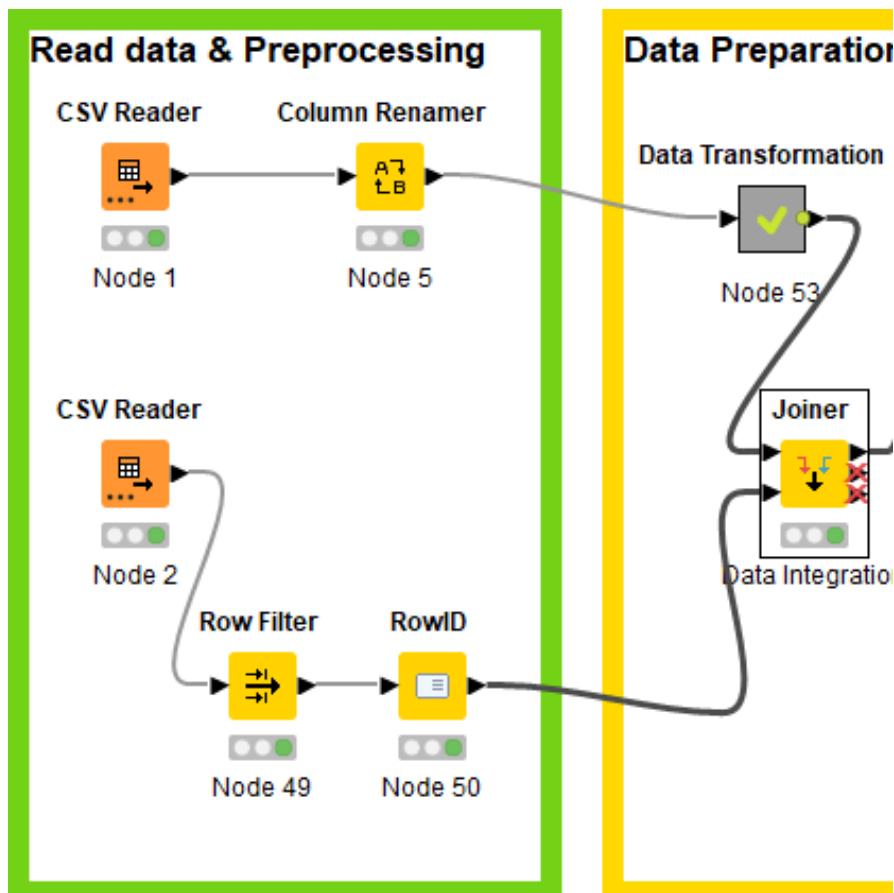
Gambar 4. 29 Tampilan kolom data setelah dikonfigurasi yang pada gambar sebelumnya belum berurutan



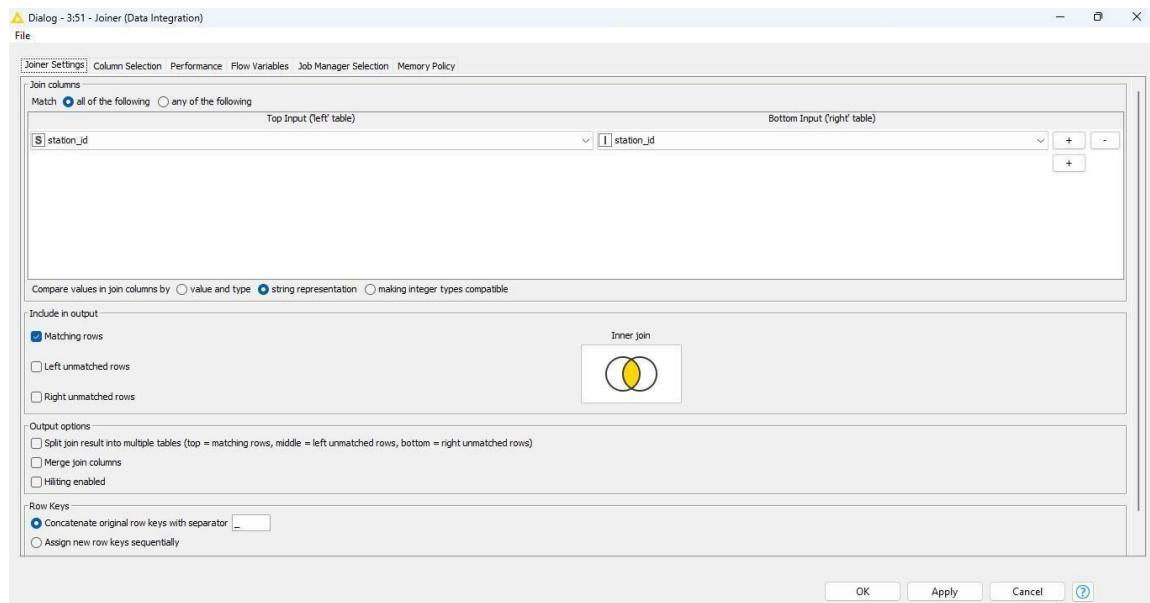
Gambar 4. 30 Konfigurasi *row id* untuk mengatur nomor baris secara berurutan

4.3.2 Data Integration

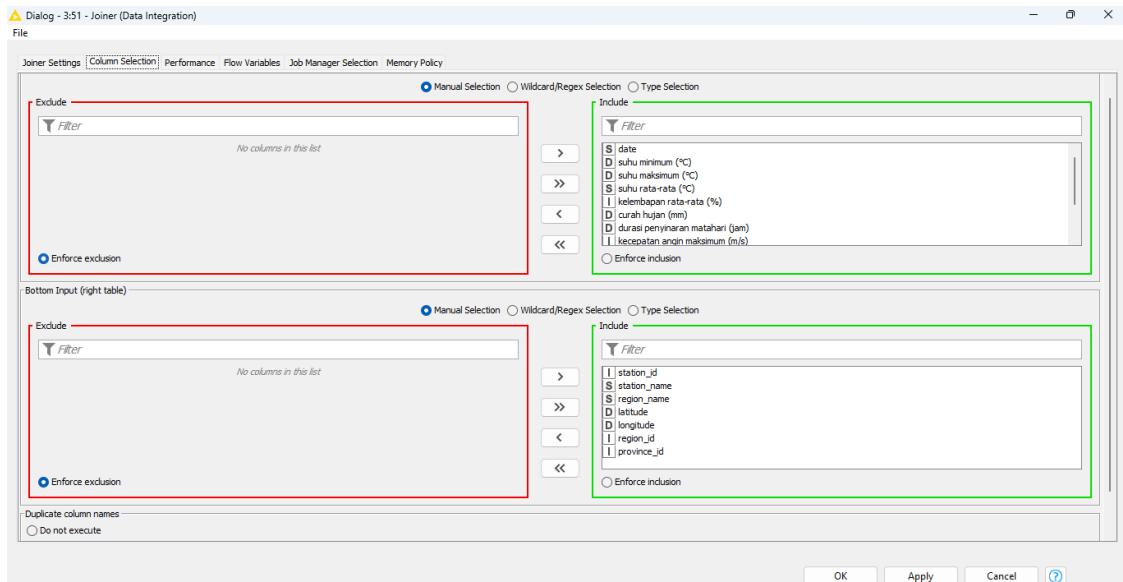
Data integration adalah penggabungan data asal berbagai dataset ke dalam satu dataset baru. Seringkali data yang diperlukan buat data mining tidak hanya berasal dari satu dataset tetapi pula dari beberapa dataset atau arsip teks. *Data integration* dilakukan di atribut-atribut yang mengidentifikasi entitas-entitas yang unik mirip *station id*, *station name*, *region name*, *latitude*, *longitude* dan lainnya. Berikut ini adalah langkah-langkah untuk tahap *data integration* pada Gambar 4. 31 sampai dengan Gambar 4. 34 dibawah ini:



Gambar 4. 31 Node joiner untuk tahap *data integration*



Gambar 4. 32 Konfigurasi node joiner tab *joiner settings*



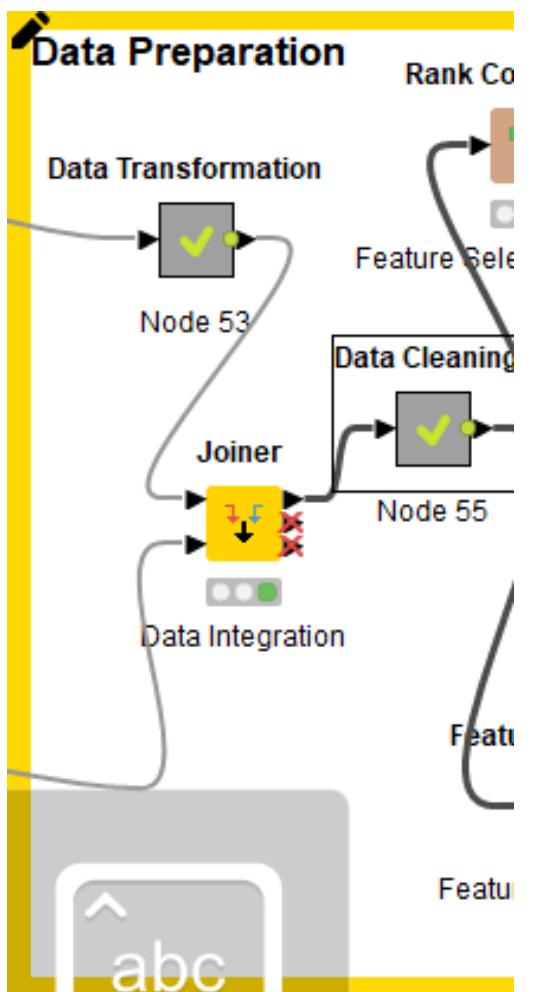
Gambar 4. 33 Konfigurasi node joiner tab column selection

Row ID	I	kelemb...	D	curah h...	D	durasi ...	I	kecepa...	I	arah an...	I	kecepa...	S	arah an...	S	station_id	I	station...	S	station_name	S	region_name	D	latitude	D	longitude	I	region_id	I	provinc...
Row0_Row0	31	1.5	4.9	20	330	4	NW	96297	96297	Stasiun Geofisika Lampung Utara	Kab. Lampung Utara	-4.836	104.87	126	8															
Row1_Row0	33	?	7.5	2	340	1	N	96297	96297	Stasiun Geofisika Lampung Utara	Kab. Lampung Utara	-4.836	104.87	126	8															
Row2_Row0	39	38	4.2	2	240	1	N	96297	96297	Stasiun Geofisika Lampung Utara	Kab. Lampung Utara	-4.836	104.87	126	8															
Row3_Row0	39	12	1.8	2	340	0	N	96297	96297	Stasiun Geofisika Lampung Utara	Kab. Lampung Utara	-4.836	104.87	126	8															
Row4_Row0	37	?	3.5	0	0	0	N	96297	96297	Stasiun Geofisika Lampung Utara	Kab. Lampung Utara	-4.836	104.87	126	8															
Row5_Row0	37	19	2	6	310	1	N	96297	96297	Stasiun Geofisika Lampung Utara	Kab. Lampung Utara	-4.836	104.87	126	8															
Row6_Row0	93	63	5.5	1	300	0	N	96297	96297	Stasiun Geofisika Lampung Utara	Kab. Lampung Utara	-4.836	104.87	126	8															
Row7_Row0	39	25	0	1	330	0	N	96297	96297	Stasiun Geofisika Lampung Utara	Kab. Lampung Utara	-4.836	104.87	126	8															
Row8_Row0	39	28	2.5	3	240	0	N	96297	96297	Stasiun Geofisika Lampung Utara	Kab. Lampung Utara	-4.836	104.87	126	8															
Row9_Row0	92	18	3	1	330	0	N	96297	96297	Stasiun Geofisika Lampung Utara	Kab. Lampung Utara	-4.836	104.87	126	8															
Row10_Row0	39	5	0	9	260	1	N	96297	96297	Stasiun Geofisika Lampung Utara	Kab. Lampung Utara	-4.836	104.87	126	8															
Row11_Row0	?	9.4	1.6	1	340	0	N	96297	96297	Stasiun Geofisika Lampung Utara	Kab. Lampung Utara	-4.836	104.87	126	8															
Row12_Row0	34	49	0	1	60	0	N	96297	96297	Stasiun Geofisika Lampung Utara	Kab. Lampung Utara	-4.836	104.87	126	8															
Row13_Row0	90	?	6.4	2	230	0	N	96297	96297	Stasiun Geofisika Lampung Utara	Kab. Lampung Utara	-4.836	104.87	126	8															

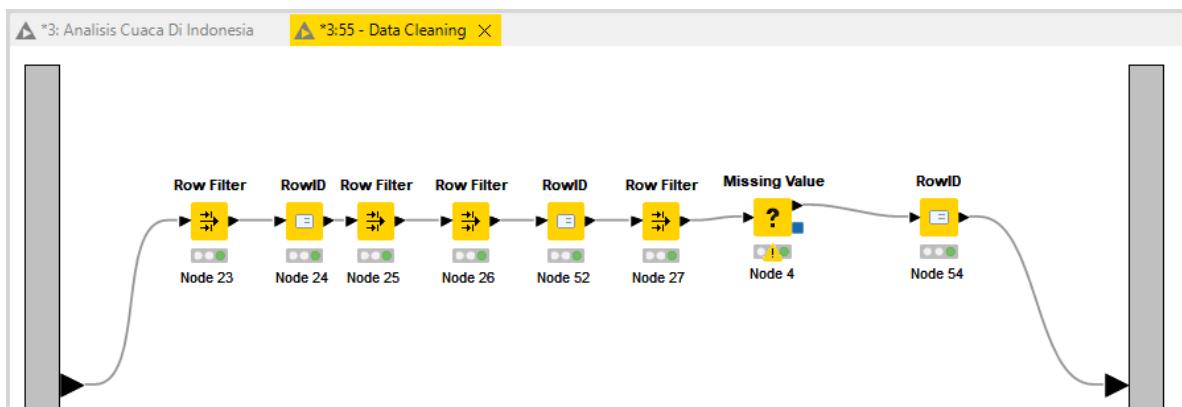
Gambar 4. 34 Tampilan gabungan 2 data

4.3.3 Data Cleaning

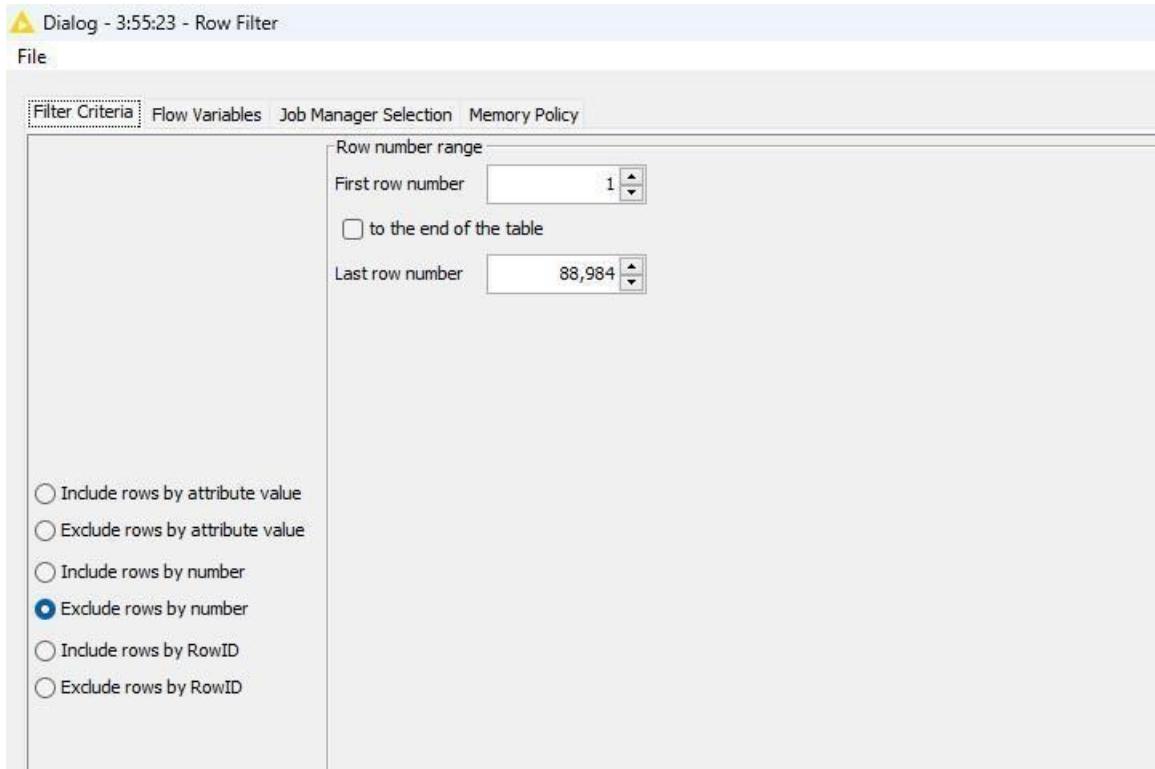
Pada tahap *data cleaning* memiliki fungsi yaitu membuang duplikasi data serta data yang tak perlu. Kemudian menyelidiki data yang tidak konsisten serta memeriksa dan membuang data yang hilang atau *missing value*. Sehingga data tadi siap buat dilakukan proses selanjutnya yaitu pemilahan data[27]. Berikut ini adalah langkah-langkah untuk proses pembersihan data atau *data cleaning* yang dipaparkan pada Gambar 4. 35 sampai dengan Gambar 4. 51 dibawah ini:



Gambar 4. 35 Tahapan *data cleaning*



Gambar 4. 36 Isi dari *metanode data cleaning*



Gambar 4. 37 Konfigurasi untuk menghapus baris 1 - 88.984

Row ID	S date	D suhu m...	D suhu m...	S suhu r...	I kelemb...	D curah h...	D durasi ...	I kecepa...	I arah an...	I kecepa...	S arah an...	S station_id	I station...	S station_name	S region_name	
Row88990_Row25	7-1-2010	24	34.8	> 25	76	0	5	3	350	2	N	96733	96733	Stasun Klimatologi Banten	Kota Tangerang Selatan	
Row88991_Row25	8-1-2010	24	29.4	> 25	86	61	0.5	3	350	0	N	96733	96733	Stasun Klimatologi Banten	Kota Tangerang Selatan	
Row88992_Row25	9-1-2010	24	31.4	> 25	84	14	0.7	3	360	2	N	96733	96733	Stasun Klimatologi Banten	Kota Tangerang Selatan	
Row88993_Row25	10-1-2010	24	31.8	> 25	84	21	1.8	3	350	2	N	96733	96733	Stasun Klimatologi Banten	Kota Tangerang Selatan	
Row88994_Row25	11-1-2010	24	32	> 25	87	0	1.9	2	270	2	W	96733	96733	Stasun Klimatologi Banten	Kota Tangerang Selatan	
Row88995_Row25	12-1-2010	24	29.8	> 25	91	5	1	4	270	3	W	96733	96733	Stasun Klimatologi Banten	Kota Tangerang Selatan	
Row88996_Row25	13-1-2010	24	29.8	> 25	87	1	1	3	270	2	W	96733	96733	Stasun Klimatologi Banten	Kota Tangerang Selatan	
Row88997_Row25	14-1-2010	23	29.4	> 25	96	8	2.5	3	350	2	N	96733	96733	Stasun Klimatologi Banten	Kota Tangerang Selatan	
Row88998_Row25	15-1-2010	21	31	< 25	88	42.7	1.7	2	360	1	N	96733	96733	Stasun Klimatologi Banten	Kota Tangerang Selatan	
Row88999_Row25	16-1-2010	22	28.6	< 25	88	7	0	2	270	0	N	96733	96733	Stasun Klimatologi Banten	Kota Tangerang Selatan	
Row89000_Row25	17-1-2010	24	29.4	> 25	91	7	0	4	270	2	W	96733	96733	Stasun Klimatologi Banten	Kota Tangerang Selatan	
Row89001_Row25	18-1-2010	23	32.8	> 25	95	35	6.9	2.6	3	360	2	N	96733	96733	Stasun Klimatologi Banten	Kota Tangerang Selatan
Row89002_Row25	19-1-2010	23	30	> 25	99	29	20.3	1	360	2	N	96733	96733	Stasun Klimatologi Banten	Kota Tangerang Selatan	
Row89003_Row25	20-1-2010	24	30.8	> 25	90	3.2	0.2	5	270	2	W	96733	96733	Stasun Klimatologi Banten	Kota Tangerang Selatan	
Row89004_Row25	21-1-2010	24	33.2	> 25	80	10	4.3	3	360	2	N	96733	96733	Stasun Klimatologi Banten	Kota Tangerang Selatan	
Row89005_Row25	22-1-2010	24	31.2	> 25	89	0	2.2	3	270	2	W	96733	96733	Stasun Klimatologi Banten	Kota Tangerang Selatan	
Row89006_Row25	23-1-2010	24	33.4	> 25	81	9.4	5.8	3	270	2	W	96733	96733	Stasun Klimatologi Banten	Kota Tangerang Selatan	
Row89007_Row25	24-1-2010	25	34	> 25	84	0	7.1	3	350	2	N	96733	96733	Stasun Klimatologi Banten	Kota Tangerang Selatan	
Row89008_Row25	25-1-2010	24	31.4	> 25	84	4	1.8	3	360	2	N	96733	96733	Stasun Klimatologi Banten	Kota Tangerang Selatan	
Row89009_Row25	26-1-2010	24	31.2	> 25	86	5.2	1	3	360	2	N	96733	96733	Stasun Klimatologi Banten	Kota Tangerang Selatan	
Row89010_Row25	27-1-2010	24	34	> 25	84	2.2	6.6	3	270	2	W	96733	96733	Stasun Klimatologi Banten	Kota Tangerang Selatan	
Row89011_Row25	28-1-2010	24	33.6	> 25	78	3.2	5.8	3	360	2	N	96733	96733	Stasun Klimatologi Banten	Kota Tangerang Selatan	
Row89012_Row25	29-1-2010	25	33.2	> 25	82	0	6.3	4	360	2	N	96733	96733	Stasun Klimatologi Banten	Kota Tangerang Selatan	
Row89013_Row25	30-1-2010	25	33	> 25	80	9.6	6.4	3	360	2	N	96733	96733	Stasun Klimatologi Banten	Kota Tangerang Selatan	
Row89014_Row25	31-1-2010	24	31	> 25	88	15.5	0	1	360	0	N	96733	96733	Stasun Klimatologi Banten	Kota Tangerang Selatan	
Row89015_Row25	1-2-2010	24	33.6	> 25	82	0.1	3.3	3	360	2	N	96733	96733	Stasun Klimatologi Banten	Kota Tangerang Selatan	
Row89016_Row25	2-2-2010	24	33.8	> 25	83	19	4.1	4	270	2	W	96733	96733	Stasun Klimatologi Banten	Kota Tangerang Selatan	
Row89017_Row25	3-2-2010	25	33.2	> 25	85	0	3.1	3	360	2	N	96733	96733	Stasun Klimatologi Banten	Kota Tangerang Selatan	
Row89018_Row25	4-2-2010	24	33	> 25	82	0.8	4.9	2	90	0	N	96733	96733	Stasun Klimatologi Banten	Kota Tangerang Selatan	
Row89019_Row25	5-2-2010	25	32	> 25	85	1.6	2.7	3	360	2	N	96733	96733	Stasun Klimatologi Banten	Kota Tangerang Selatan	
Row89020_Row25	6-2-2010	25	33	> 25	86	0.5	3	3	360	2	N	96733	96733	Stasun Klimatologi Banten	Kota Tangerang Selatan	
Row89021_Row25	7-2-2010	24	33.8	> 25	80	3	6.5	2	360	2	N	96733	96733	Stasun Klimatologi Banten	Kota Tangerang Selatan	
Row89022_Row25	8-2-2010	25	33.8	> 25	80	0	3.3	3	350	2	N	96733	96733	Stasun Klimatologi Banten	Kota Tangerang Selatan	
Row89023_Row25	9-2-2010	25	35	> 25	75	0	6.2	2	360	2	N	96733	96733	Stasun Klimatologi Banten	Kota Tangerang Selatan	
Row89024_Row25	10-2-2010	24	34.2	> 25	78	0.5	5.7	3	360	2	N	96733	96733	Stasun Klimatologi Banten	Kota Tangerang Selatan	
Row89025_Row25	11-2-2010	25	33.8	> 25	81	0.5	4.6	4	360	2	N	96733	96733	Stasun Klimatologi Banten	Kota Tangerang Selatan	
Row89026_Row25	12-2-2010	25	34	> 25	77	0	4.6	2	350	2	N	96733	96733	Stasun Klimatologi Banten	Kota Tangerang Selatan	
Row89027_Row25	13-2-2010	25	30	> 25	93	0.5	0	2	270	0	N	96733	96733	Stasun Klimatologi Banten	Kota Tangerang Selatan	

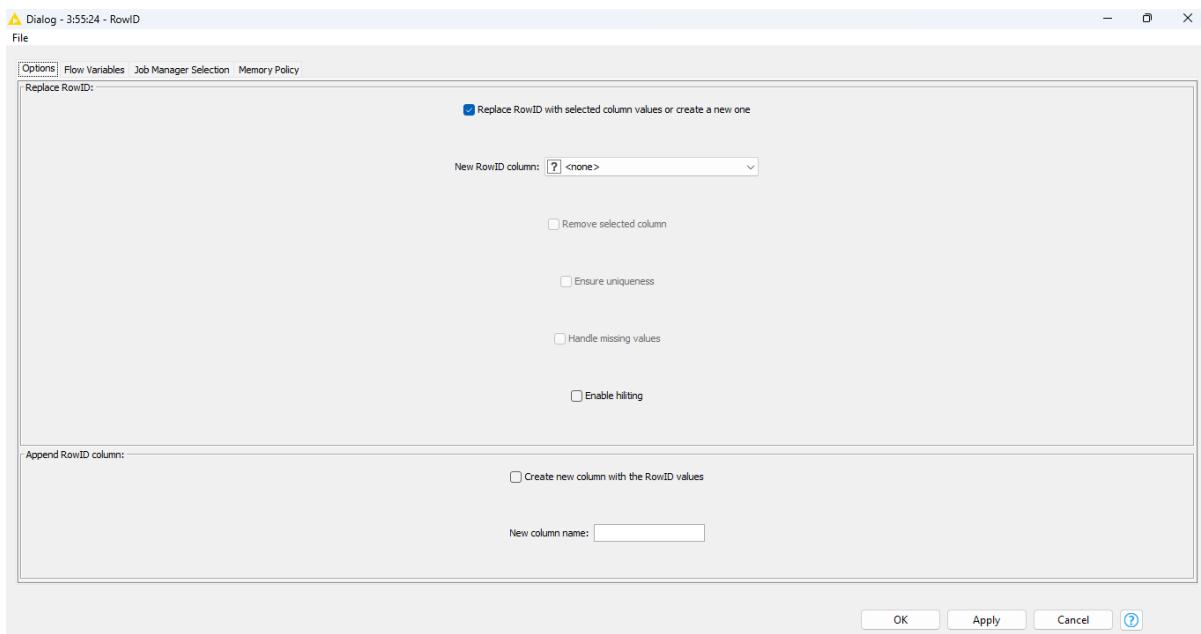
Gambar 4. 38 Tampilan tabel data setelah di konfigurasi node row filter

Penjelasan pada Gambar 4. 37 dan Gambar 4. 38 yaitu proses pembuangan data duplikat dan data yang tidak perlu. Baris data yang dihapus tersebut berisi data selain wilayah Jakarta, Bogor, dan Tangerang. Diantara baris data yang dihapus berdasarkan

row id yakni Kabupaten Lampung Utara, Kabupaten Nunukan, Kota Tarakan, Kabupaten Bulungan, Kabupaten Berau, Kabupaten Sambas, Kabupaten Melawi, Kabupaten Sintang, Kabupaten Kapuas Hulu, Kabupaten Kubu Raya, Kabupaten/Kota Pontianak, Kabupaten Barito Utara, Kota Samarinda, Kabupaten Ketapang, Kota Balikpapan, Kabupaten Kotawaringin Barat, Kabupaten Kotawaringin Timur, Kabupaten Barito Selatan, Kota Palangkaraya, Kota Banjarmasin, Kota Banjarbaru, dan Kabupaten Kotabaru sebagaimana tampilan tabel data sebelum dikonfigurasi yang ditampilkan pada Gambar 4. 39 dibawah ini:

Row88976_Row24	88	6.4	1.5	4	270	1	C	96695	96695	Stasiun Meteorologi Gusti Sya...	Kab. Kotabaru	-3.3	116.17	339	22
Row88977_Row24	?	1.2	4.6	4	90	1	C	96695	96695	Stasiun Meteorologi Gusti Sya...	Kab. Kotabaru	-3.3	116.17	339	22
Row88978_Row24	33	28.4	3	4	250	1	C	96695	96695	Stasiun Meteorologi Gusti Sya...	Kab. Kotabaru	-3.3	116.17	339	22
Row88979_Row24	33	0	3.8	4	180	1	C	96695	96695	Stasiun Meteorologi Gusti Sya...	Kab. Kotabaru	-3.3	116.17	339	22
Row88980_Row24	31	0.2	?	6	150	2	SE	96695	96695	Stasiun Meteorologi Gusti Sya...	Kab. Kotabaru	-3.3	116.17	339	22
Row88981_Row24	38	22.7	5	3	150	1	C	96695	96695	Stasiun Meteorologi Gusti Sya...	Kab. Kotabaru	-3.3	116.17	339	22
Row88982_Row24	?	5.4	0	6	170	2	SE	96695	96695	Stasiun Meteorologi Gusti Sya...	Kab. Kotabaru	-3.3	116.17	339	22
Row88983_Row24	32	1.6	5.8	5	170	2	S	96695	96695	Stasiun Meteorologi Gusti Sya...	Kab. Kotabaru	-3.3	116.17	339	22
Row88984_Row25	78	0.1	4.8	3	270	2	W	96733	96733	Stasiun Klimatologi Banten	Kota Tangerang Selata	-6.262	106.751	269	16
Row88985_Row25	78	0	5.9	4	270	2	W	96733	96733	Stasiun Klimatologi Banten	Kota Tangerang Selata	-6.262	106.751	269	16
Row88986_Row25	79	14.6	4.2	4	270	3	W	96733	96733	Stasiun Klimatologi Banten	Kota Tangerang Selata	-6.262	106.751	269	16
Row88987_Row25	72	0	8	5	270	4	W	96733	96733	Stasiun Klimatologi Banten	Kota Tangerang Selata	-6.262	106.751	269	16
Row88988_Row25	77	0	1.7	4	225	3	SW	96733	96733	Stasiun Klimatologi Banten	Kota Tangerang Selata	-6.262	106.751	269	16
Row88989_Row25	74	?	5.4	5	270	3	W	96733	96733	Stasiun Klimatologi Banten	Kota Tangerang Selata	-6.262	106.751	269	16
Row88990_Row25	76	0	5	3	350	2	N	96733	96733	Stasiun Klimatologi Banten	Kota Tangerang Selata	-6.262	106.751	269	16
Row88991_Row25	86	61	0.5	3	350	0	N	96733	96733	Stasiun Klimatologi Banten	Kota Tangerang Selata	-6.262	106.751	269	16

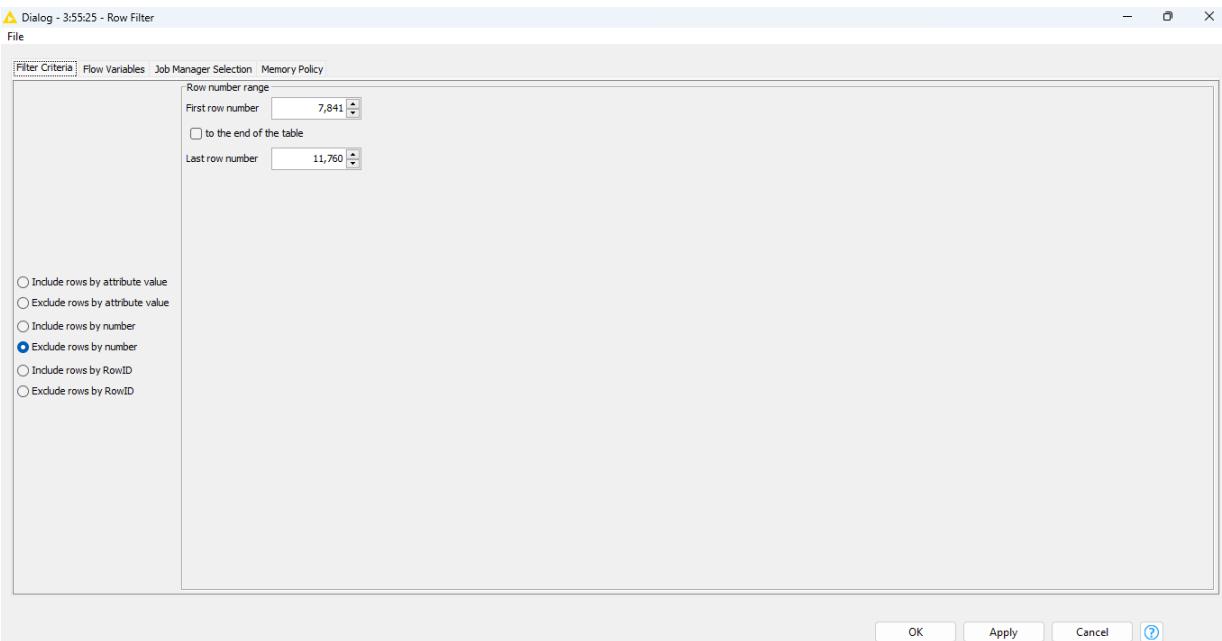
Gambar 4. 39 Tampilan tabel data sebelum dikonfigurasi



Gambar 4. 40 Konfigurasi untuk menetapkan kembali ID baris

Row ID	S date	D suhu minim...	D suhu m...	S suhu ra...	I kelemb...	D curah h...	D durasi ...	I kecepa...	I arah an...	I kecepa...	S arah an...	S station_id	I station...	S station_name	S region_name
Row0	1-1-2010	25	33.4	> 25	78	0.1	4.8	3	270	2	W	96733	96733	Stasiun Klimatologi Banten	Kota Tangerang Selata
Row1	2-1-2010	25	34.4	> 25	78	0	5.9	4	270	2	W	96733	96733	Stasiun Klimatologi Banten	Kota Tangerang Selata
Row2	3-1-2010	25	33.4	> 25	79	14.6	4.2	4	270	3	W	96733	96733	Stasiun Klimatologi Banten	Kota Tangerang Selata
Row3	4-1-2010	26	34.4	> 25	72	0	8	5	270	4	W	96733	96733	Stasiun Klimatologi Banten	Kota Tangerang Selata
Row4	5-1-2010	26	32.8	> 25	77	0	1.7	4	225	3	SW	96733	96733	Stasiun Klimatologi Banten	Kota Tangerang Selata
Row5	6-1-2010	26	33.4	> 25	74	?	5.4	5	270	3	W	96733	96733	Stasiun Klimatologi Banten	Kota Tangerang Selata
Row6	7-1-2010	24	34.8	> 25	76	0	5	3	350	2	N	96733	96733	Stasiun Klimatologi Banten	Kota Tangerang Selata
Row7	8-1-2010	24	29.4	> 25	86	61	0.5	3	350	0	N	96733	96733	Stasiun Klimatologi Banten	Kota Tangerang Selata
Row8	9-1-2010	24	31.4	> 25	84	14	0.7	3	360	2	N	96733	96733	Stasiun Klimatologi Banten	Kota Tangerang Selata
Row9	10-1-2010	24	31.8	> 25	84	21	1.8	3	350	2	N	96733	96733	Stasiun Klimatologi Banten	Kota Tangerang Selata
Row10	11-1-2010	24	32	> 25	87	0	1.9	2	270	2	W	96733	96733	Stasiun Klimatologi Banten	Kota Tangerang Selata
Row11	12-1-2010	24	29.8	> 25	91	5	1	4	270	3	W	96733	96733	Stasiun Klimatologi Banten	Kota Tangerang Selata
Row12	13-01-2010	24	29.8	> 25	87	1	1	3	270	2	W	96733	96733	Stasiun Klimatologi Banten	Kota Tangerang Selata
Row13	14-01-2010	23	29.4	> 25	96	8	2.5	3	350	2	N	96733	96733	Stasiun Klimatologi Banten	Kota Tangerang Selata
Row14	15-01-2010	21	31	< 25	88	42.7	1.7	2	360	1	N	96733	96733	Stasiun Klimatologi Banten	Kota Tangerang Selata
Row15	16-01-2010	22	28.6	< 25	88	?	0	2	270	0	N	96733	96733	Stasiun Klimatologi Banten	Kota Tangerang Selata
Row16	17-01-2010	24	29.4	> 25	91	7	0	4	270	2	W	96733	96733	Stasiun Klimatologi Banten	Kota Tangerang Selata
Row17	18-01-2010	23	32.8	> 25	85	6.9	2.6	3	360	2	N	96733	96733	Stasiun Klimatologi Banten	Kota Tangerang Selata
Row18	19-01-2010	23	30	> 25	89	20.3	1	3	360	2	N	96733	96733	Stasiun Klimatologi Banten	Kota Tangerang Selata
Row19	20-01-2010	24	30.8	> 25	90	3.2	0.2	5	270	2	W	96733	96733	Stasiun Klimatologi Banten	Kota Tangerang Selata
Row20	21-01-2010	24	33.2	> 25	80	10	4.3	3	360	2	N	96733	96733	Stasiun Klimatologi Banten	Kota Tangerang Selata
Row21	22-01-2010	24	31.2	> 25	89	0	2.2	3	270	2	W	96733	96733	Stasiun Klimatologi Banten	Kota Tangerang Selata
Row22	23-01-2010	24	33.4	> 25	81	9.4	5.8	3	270	2	W	96733	96733	Stasiun Klimatologi Banten	Kota Tangerang Selata
Row23	24-01-2010	25	34	> 25	84	0	7.1	3	350	2	N	96733	96733	Stasiun Klimatologi Banten	Kota Tangerang Selata
Row24	25-01-2010	24	31.4	> 25	84	4	1.8	3	360	2	N	96733	96733	Stasiun Klimatologi Banten	Kota Tangerang Selata
Row25	26-01-2010	24	31.2	> 25	86	5.2	1	3	360	2	N	96733	96733	Stasiun Klimatologi Banten	Kota Tangerang Selata
Row26	27-01-2010	24	34	> 25	84	2.2	6.6	3	270	2	W	96733	96733	Stasiun Klimatologi Banten	Kota Tangerang Selata
Row27	28-01-2010	24	33.6	> 25	78	3.2	5.8	3	360	2	N	96733	96733	Stasiun Klimatologi Banten	Kota Tangerang Selata
Row28	29-01-2010	25	33.2	> 25	82	0	6.3	4	360	2	N	96733	96733	Stasiun Klimatologi Banten	Kota Tangerang Selata
Row29	30-01-2010	25	33	> 25	80	9.6	6.4	3	360	2	N	96733	96733	Stasiun Klimatologi Banten	Kota Tangerang Selata
Row30	31-01-2010	24	31	> 25	88	15.5	0	1	360	0	N	96733	96733	Stasiun Klimatologi Banten	Kota Tangerang Selata
Row31	1-2-2010	24	33.6	> 25	82	0.1	3.3	3	360	2	N	96733	96733	Stasiun Klimatologi Banten	Kota Tangerang Selata
Row32	2-2-2010	24	33.8	> 25	83	19	4.1	4	270	2	W	96733	96733	Stasiun Klimatologi Banten	Kota Tangerang Selata
Row33	3-2-2010	25	33.2	> 25	85	0	3.1	3	360	2	N	96733	96733	Stasiun Klimatologi Banten	Kota Tangerang Selata
Row34	4-2-2010	24	33	> 25	82	0.8	4.9	2	90	0	N	96733	96733	Stasiun Klimatologi Banten	Kota Tangerang Selata
Row35	5-2-2010	25	32	> 25	85	1.6	2.7	3	360	2	N	96733	96733	Stasiun Klimatologi Banten	Kota Tangerang Selata
Row36	6-2-2010	25	33	> 25	86	0.5	3	3	360	2	N	96733	96733	Stasiun Klimatologi Banten	Kota Tangerang Selata
Row37	7-2-2010	24	33.8	> 25	80	3	6.5	2	360	2	N	96733	96733	Stasiun Klimatologi Banten	Kota Tangerang Selata

Gambar 4. 41 Tampilan tabel data setelah konfigurasi *row id*

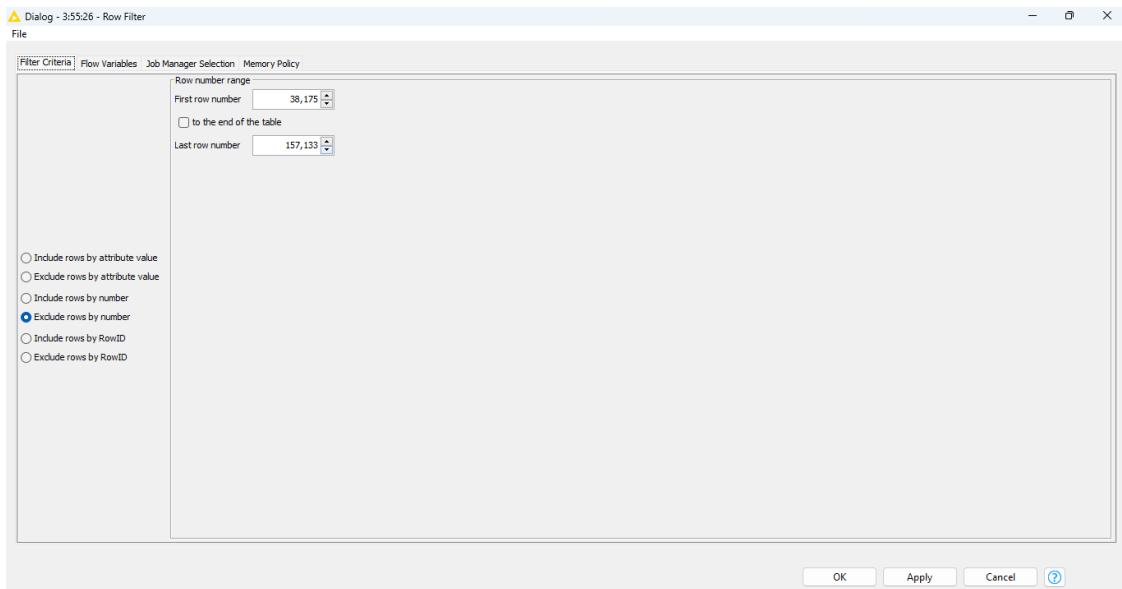


Gambar 4. 42 Menghapus baris 7.841 - 11.760

Penjelasan pada Gambar 4. 42 yaitu proses pembuangan data yang tidak perlu. Baris data yang dihapus tersebut berisi data selain wilayah Jakarta, Bogor, dan Tangerang. Diantara baris data yang dihapus berdasarkan *row id* yakni Kota Serang. Sebagaimana tampilan tabel data setelah dikonfigurasi menggunakan *node row filter* yang dipaparkan pada Gambar 4. 43 dibawah ini:

Row7836	28-12-2020	25	33	> 25	88	?	8.7	5	270	3	W	96735	96735	Stasiun Geofisika Tange... Kota Tangerang
Row7837	29-12-2020	24.6	31.2	> 25	89	?	3.6	6	270	3	W	96735	96735	Stasiun Geofisika Tange... Kota Tangerang
Row7838	30-12-2020	24.8	31	> 25	88	?	1.1	4	270	2	W	96735	96735	Stasiun Geofisika Tange... Kota Tangerang
Row7839	31-12-2020	24.6	29.8	> 25	21	14	1.2	3	259	1	N	96735	96735	Stasiun Geofisika Tange... Kota Tangerang
Row11760	1-1-2010	24	32.8	> 25	78	0	3.8	8	225	3	W	96739	96739	Stasiun Meteorologi Budi... Kab. Tangerang
Row11761	2-1-2010	25	33.6	> 25	77	4	6	7	270	4	W	96739	96739	Stasiun Meteorologi Budi... Kab. Tangerang
Row11762	3-1-2010	25	32.2	> 25	76	0	2.5	7	270	5	W	96739	96739	Stasiun Meteorologi Budi... Kab. Tangerang
Row11763	4-1-2010	25	33.8	> 25	76	0	7	10	270	6	W	96739	96739	Stasiun Meteorologi Budi... Kab. Tangerang
Row11764	5-1-2010	25	31.4	> 25	74	?	1.1	8	270	5	W	96739	96739	Stasiun Meteorologi Budi... Kab. Tangerang
Row11765	6-1-2010	25	33.6	> 25	78	0	6.7	9	270	5	W	96739	96739	Stasiun Meteorologi Budi... Kab. Tangerang
Row11766	7-1-2010	24	33.4	> 25	73	?	6.6	7	225	4	SW	96739	96739	Stasiun Meteorologi Budi... Kab. Tangerang
Row11767	8-1-2010	23	29	> 25	89	1.5	0.6	4	225	3	N	96739	96739	Stasiun Meteorologi Budi... Kab. Tangerang
Row11768	9-1-2010	23	30.3	< 25	91	22.5	1	5	315	2	S	96739	96739	Stasiun Meteorologi Budi... Kab. Tangerang
Row11769	10-1-2010	24	30.4	> 25	86	2	0	2	225	2	W	96739	96739	Stasiun Meteorologi Budi... Kab. Tangerang
Row11770	11-1-2010	24	31.6	> 25	84	2	2.2	5	270	3	W	96739	96739	Stasiun Meteorologi Budi... Kab. Tangerang
Row11771	12-1-2010	23	29	> 25	85	11	0	7	270	4	W	96739	96739	Stasiun Meteorologi Budi... Kab. Tangerang
Row11772	13-1-2010	23	29.8	> 25	88	34.1	0.6	8	315	4	W	96739	96739	Stasiun Meteorologi Budi... Kab. Tangerang
Row11773	14-01-2010	21	28.3	< 25	86	23.4	0.6	6	270	3	W	96739	96739	Stasiun Meteorologi Budi... Kab. Tangerang
Row11774	15-01-2010	22	28.8	< 25	85	2	0	3	360	3	W	96739	96739	Stasiun Meteorologi Budi... Kab. Tangerang
Row11775	16-01-2010	23	28.8	< 25	90	5	0	4	315	3	SW	96739	96739	Stasiun Meteorologi Budi... Kab. Tangerang
Row11776	17-01-2010	23	28	> 25	91	16.5	0	4	270	3	W	96739	96739	Stasiun Meteorologi Budi... Kab. Tangerang
Row11777	18-01-2010	24	30.8	> 25	84	9.1	2.6	4	270	2	W	96739	96739	Stasiun Meteorologi Budi... Kab. Tangerang

Gambar 4. 43 Tampilan tabel data setelah dikonfigurasi menggunakan *node row filter*



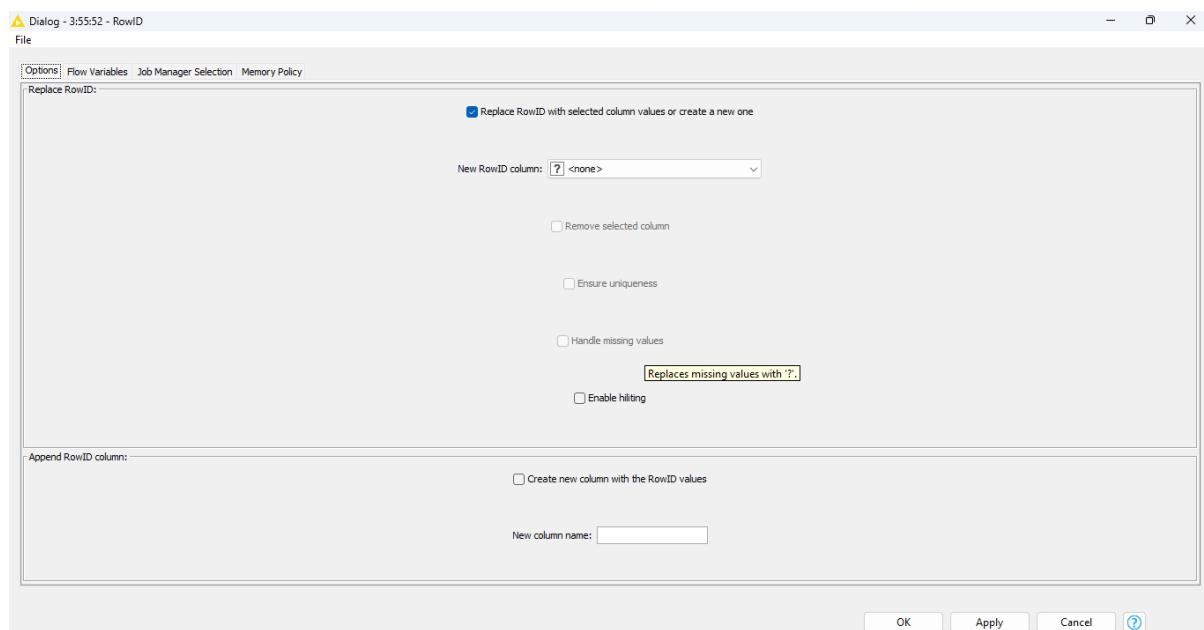
Gambar 4. 44 Menghapus baris 38.175 - 157.133

Penjelasan pada Gambar 4. 44 yaitu proses pembuangan data yang tidak perlu. Baris data yang dihapus tersebut berisi data selain wilayah Jakarta, Bogor, dan Tangerang. Diantara baris data yang dihapus berdasarkan *row id* yakni Kabupaten Majalengka, Kabupaten Cirebon, Kabupaten Tegal, Kabupaten Cilacap, Kabupaten Banjarnegeara, Kota Semarang, Kabupaten Sleman, Kabupaten Gresik, Kota Surabaya, Kabupaten Sidoarjo, Kabupaten Tuban, Kabupaten Malang, Kabupaten Pasuruan, Kabupaten Jayawijaya, Kabupaten Jayapura, Kabupaten Manokwari, Kota Jayapura, Kabupaten Buru, Kabupaten Maluku Tengah, Kota Ambon, Kabupaten Seram Bagian Barat, Kabupaten Seram Bagian Timur, Kabupaten Kaimana, Kabupaten Paniai, Kabupaten

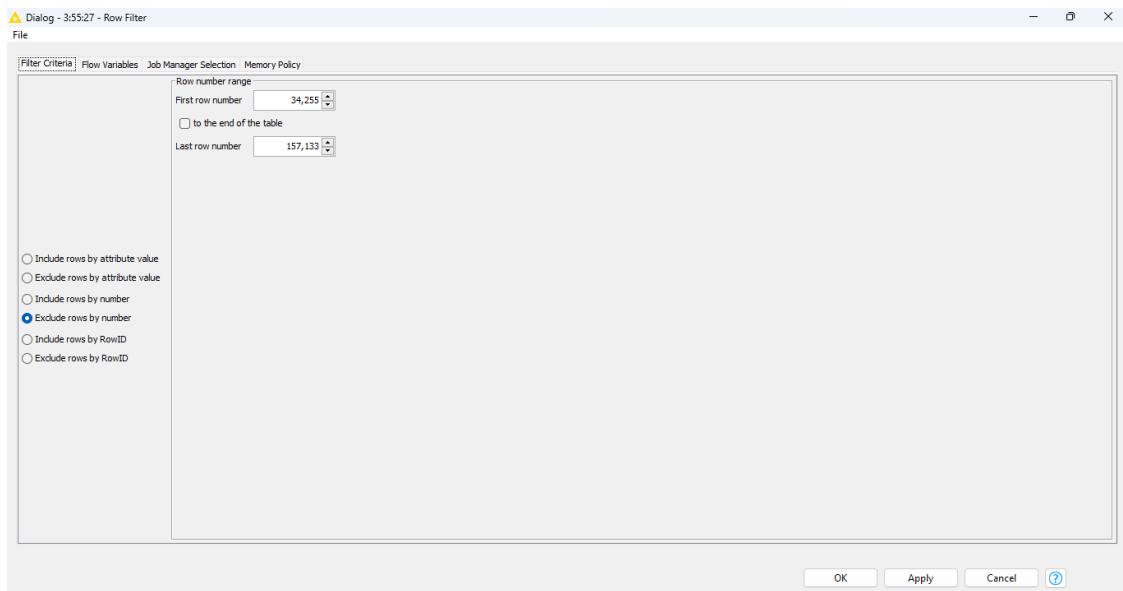
Maluku Tengah, Kabupaten Mimika, Kabupaten Maluku Tenggara, Kabupaten Boven Digoel, Kabupaten Maluku Tengara Barat, dan Kabupaten Merauke. Sebagaimana tampilan tabel data setelah dikonfigurasi menggunakan *node row filter* yang dipaparkan pada Gambar 4. 45 dibawah ini:

KodeKab	[20-12-2010]	[41-3]	[< 2b]	[> b]	[u,y]	[d,b]	[>]	[d,u]	[d]	[NW]	[B,N,S]	[B,E,S]	Stasiun Geofisika Bandung Kota Bandung
Row42094	31-12-2010	21	28.8	< 25	74	7.6	0.2	5	360	3	NW	96783	Stasiun Geofisika Bandung Kota Bandung
Row42085	1-1-2010	24	32.8	> 25	83	3.7	5.3	2	90	2	E	96791	Stasiun Meteorologi Kert..., Kab. Majalengka
Row42086	2-1-2010	25	32.6	> 25	84	0	5.2	3	270	2	W	96791	Stasiun Meteorologi Kert..., Kab. Majalengka
Row42087	3-1-2010	24	33.2	> 25	82	1.6	4.3	2	360	2	N	96791	Stasiun Meteorologi Kert..., Kab. Majalengka
Row42088	4-1-2010	25	34	> 25	84	0	7	6	270	2	N	96791	Stasiun Meteorologi Kert..., Kab. Majalengka
Row42089	5-1-2010	24	33	> 25	85	32.8	4.6	5	90	2	S	96791	Stasiun Meteorologi Kert..., Kab. Majalengka
Row42090	6-1-2010	24	31.8	> 25	87	3.5	4.5	5	270	2	W	96791	Stasiun Meteorologi Kert..., Kab. Majalengka
Row42091	7-1-2010	24	31.8	> 25	85	1	3.3	3	135	2	W	96791	Stasiun Meteorologi Kert..., Kab. Majalengka
Row42092	8-1-2010	24	30.3	> 25	84	5.8	1.8	3	360	2	N	96791	Stasiun Meteorologi Kert..., Kab. Majalengka
Row42093	9-1-2010	24	30.1	> 25	87	0	2.4	3	45	2	N	96791	Stasiun Meteorologi Kert..., Kab. Majalengka

Gambar 4. 45 tampilan tabel data setelah dikonfigurasi menggunakan *node row id*

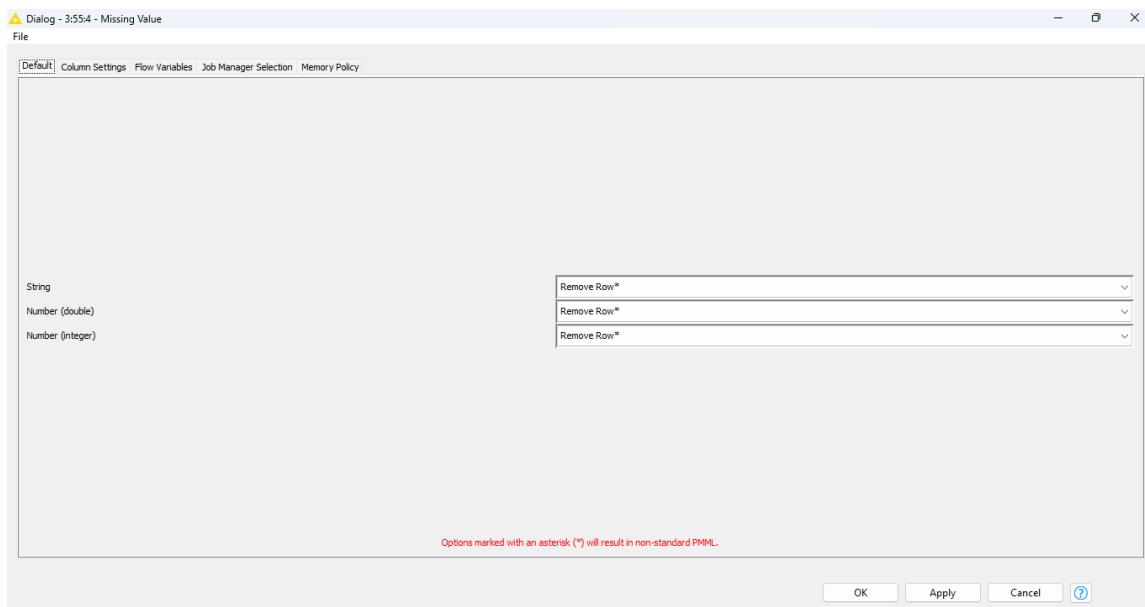


Gambar 4. 46 Konfigurasi untuk menetapkan kembali ID baris



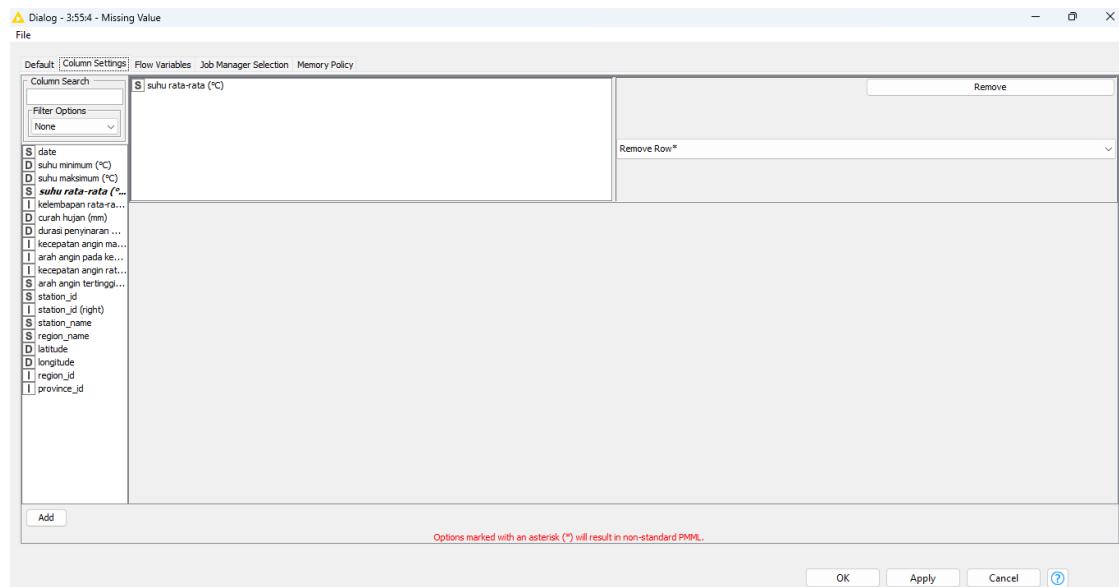
Gambar 4. 47 Konfigurasi untuk menghapus baris 34.255 - 157.133

Penjelasan pada Gambar 4. 47 yaitu proses pembuangan data yang tidak perlu. Baris data yang dihapus tersebut berisi data selain wilayah Jakarta, Bogor, dan Tangerang. Diantara baris data yang dihapus berdasarkan *row id* yakni Kota Bandung dan Kabupaten Majalengka yang masih tersisa.



Gambar 4. 48 Konfigurasi untuk menghilangkan nilai kosong menggunakan *node missing value*

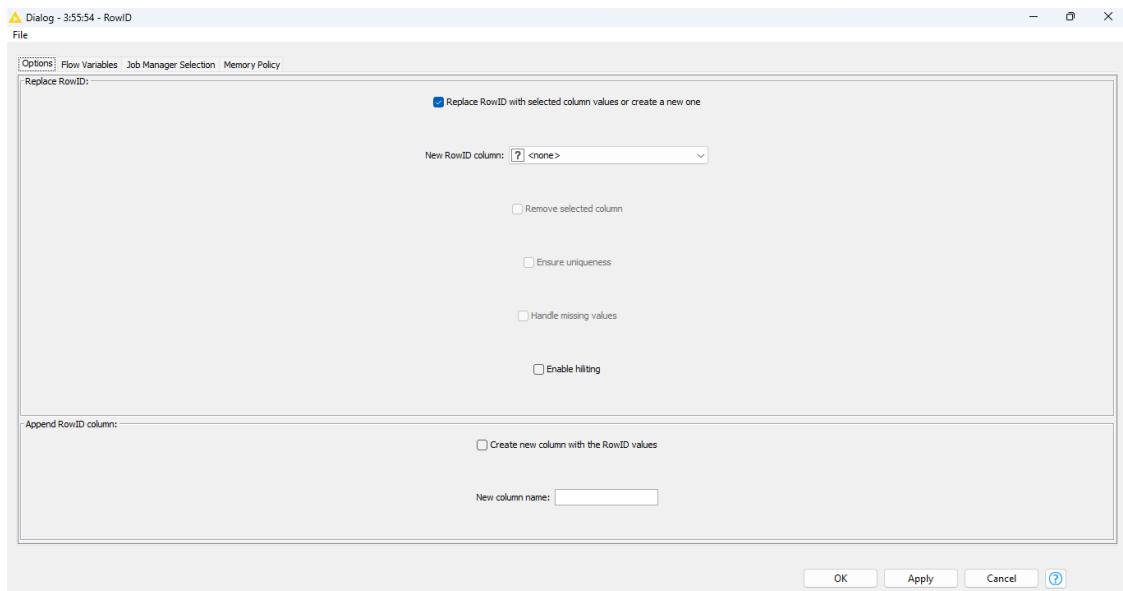
Penjelasan Gambar 4. 48 yaitu konfigurasi pada *node missing value* di tab *default* terdapat beberapa tipe data yang memiliki data yang kosong yaitu *String*, *Double*, dan *Integer*. Ketiga tipe data yang memiliki data kosong yang mana akan dihapus dengan cara mengkonfigurasikan *node missing value* lalu pilih opsi “*Remove row*”. Kemudian pada kolom yang dijadikan sebagai kategori yaitu suhu rata-rata di atur dalam tab *Column Settings* seperti pada gambar dibawah ini:



Gambar 4. 49 Konfigurasi *node missing value* tab column settings

Row	date	station_id	suhu minimum (°C)	suhu maksimum (°C)	suhu rata-rata (°C)	curah hujan (mm)	durasi penirukan ...	kecepatan angin ma...	arah angin pada ke...	kecepatan angin rat...	arah angin tertingg...	station_id	station_id (right)	station_name	region_name	latitude	longitude	region_id	province_id
Row19381	5-12-2020	26	30.2	> 25	80	14.4	4	6	270	2	C	96745	96745	Stasiun Meteorologi Kem... ,Kota Adm. Jakarta P...					
Row19382	6-12-2020	25	28.6	> 25	90	22.5	0.5	7	320	3	W	96745	96745	Stasiun Meteorologi Kem... ,Kota Adm. Jakarta P...					
Row19383	7-12-2020	24.2	30.2	> 25	88	24.8	0	3	320	1	C	96745	96745	Stasiun Meteorologi Kem... ,Kota Adm. Jakarta P...					
Row19384	8-12-2020	24.4	32.2	> 25	78	7.2	0.2	4	340	2	W	96745	96745	Stasiun Meteorologi Kem... ,Kota Adm. Jakarta P...					
Row19385	9-12-2020	25	25.6	> 25	32.3	> 25	75	9.3	6-2	10	270	2	C	96745	96745	Stasiun Meteorologi Kem... ,Kota Adm. Jakarta P...			
Row19387	11-12-2020	26	30.4	> 25	80	0.4	2	8	320	3	W	96745	96745	Stasiun Meteorologi Kem... ,Kota Adm. Jakarta P...					
Row19388	12-12-2020	25.4	32.2	> 25	78	4.8	3.3	4	300	2	W	96745	96745	Stasiun Meteorologi Kem... ,Kota Adm. Jakarta P...					
Row19392	16-12-2020	25.4	32.2	> 25	76	9	1.8	4	340	2	W	96745	96745	Stasiun Meteorologi Kem... ,Kota Adm. Jakarta P...					
Row19393	17-12-2020	24.8	31.8	> 25	79	6	2.2	5	260	2	W	96745	96745	Stasiun Meteorologi Kem... ,Kota Adm. Jakarta P...					
Row19397	21-12-2020	24.8	29.6	> 25	76	4.2	2.4	5	290	2	C	96745	96745	Stasiun Meteorologi Kem... ,Kota Adm. Jakarta P...					
Row19399	23-12-2020	25.6	34.8	> 25	65	0	2	7	5	300	2	C	96745	96745	Stasiun Meteorologi Kem... ,Kota Adm. Jakarta P...				
Row19400	24-12-2020	26.4	33.4	> 25	68	0	7	5	280	2	W	96745	96745	Stasiun Meteorologi Kem... ,Kota Adm. Jakarta P...					
Row19405	29-12-2020	26.4	32	> 25	73	0	5.3	9	290	2	C	96745	96745	Stasiun Meteorologi Kem... ,Kota Adm. Jakarta P...					
Row19407	31-12-2020	25.6	30.2	> 25	83	5	0.2	4	240	1	C	96745	96745	Stasiun Meteorologi Kem... ,Kota Adm. Jakarta P...					
Row22536	1-1-2010	24	32.3	> 25	75	0	4	11	270	4	SW	96749	96749	Stasiun Meteorologi Soe... ,Kota Tangerang					
Row22537	2-1-2010	25	33.2	> 25	81	0	5.6	16	240	4	SW	96749	96749	Stasiun Meteorologi Soe... ,Kota Tangerang					
Row22538	3-1-2010	24	32.4	> 25	75	33.7	5	11	300	5	SW	96749	96749	Stasiun Meteorologi Soe... ,Kota Tangerang					
Row22539	4-1-2010	26	33.3	> 25	74	0	8	14	260	6	SW	96749	96749	Stasiun Meteorologi Soe... ,Kota Tangerang					
Row22540	5-1-2010	25	31.4	> 25	74	0	1	13	270	5	SW	96749	96749	Stasiun Meteorologi Soe... ,Kota Tangerang					
Row22541	6-1-2010	26	32.3	> 25	78	0	6.4	11	150	5	W	96749	96749	Stasiun Meteorologi Soe... ,Kota Tangerang					
Row22542	7-1-2010	25	32.8	> 25	77	0	6	14	250	3	N	96749	96749	Stasiun Meteorologi Soe... ,Kota Tangerang					

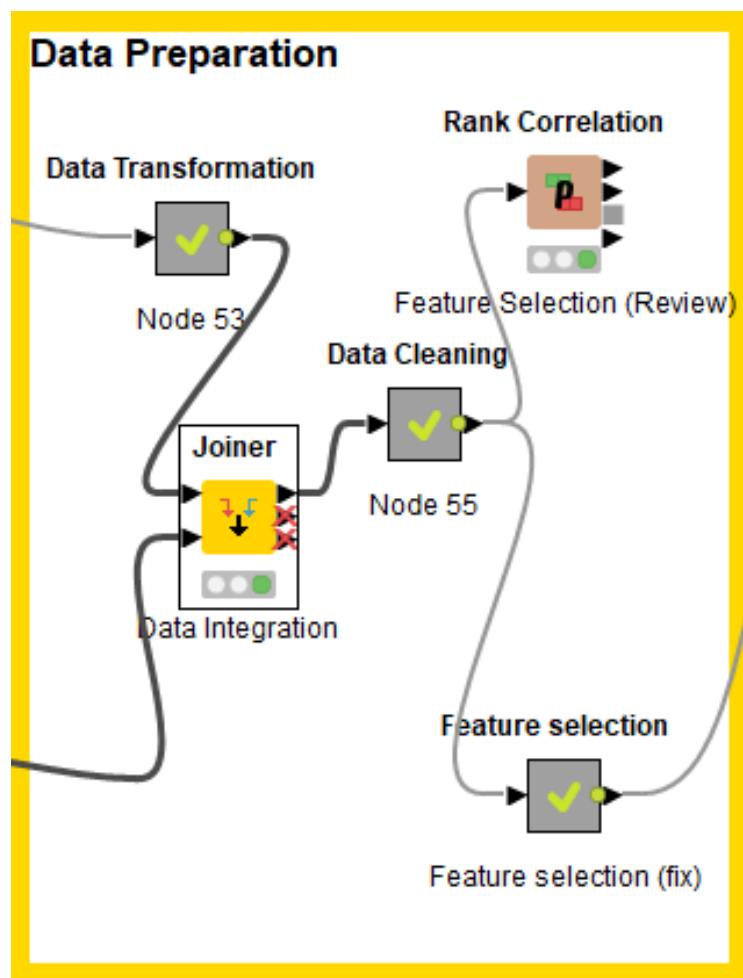
Gambar 4. 50 Tabel Data Setelah Dikonfigurasikan Menggunakan *Node Missing Value*



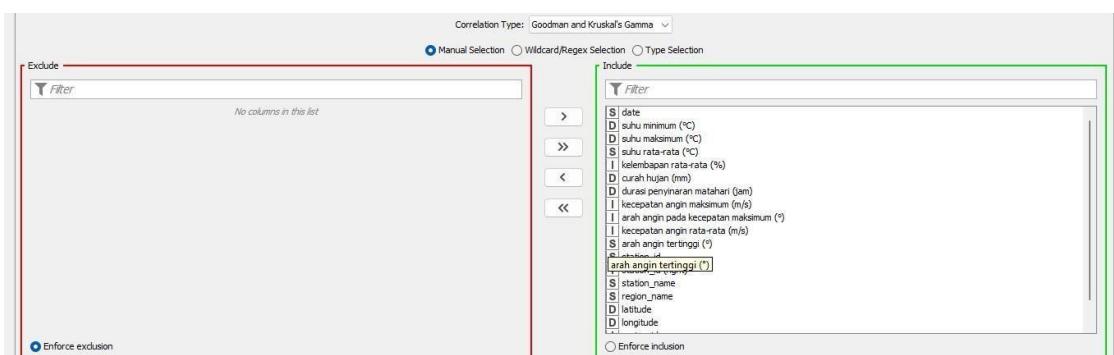
Gambar 4. 51 Konfigurasi untuk menetapkan kembali ID baris

4.3.4 Feature Selection

Pada tahap *feature selection* yaitu buat memilih fitur apa saja yang dipergunakan pada waktu pemodelan memakai metode *classification*. Proses seleksi fitur dilakukan dengan memilih fitur yang berpengaruh terhadap yang akan terjadi penjabaran menjadi lebih baik serta fitur yang dapat diproses oleh contoh *Machine Learning*. Hasil dari tahapan ini mengurangi fitur yang tadinya ada 19 kolom menjadi 8 kolom[28]. Berikut ini adalah langkah-langkah untuk tahapan feature selection dibawah ini:



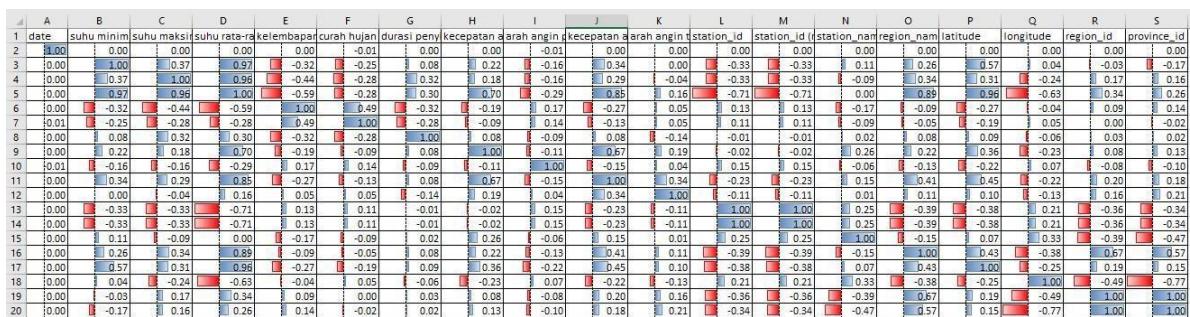
Gambar 4. 52 Metanode tahap *feature selection*



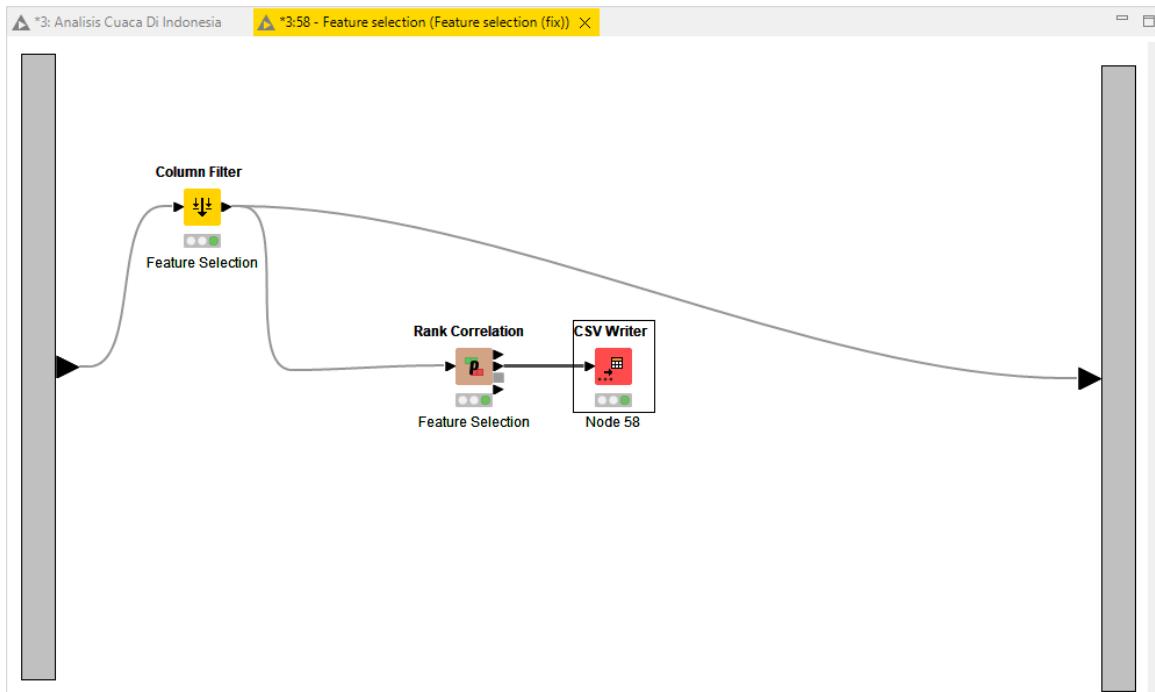
Gambar 4. 53 Konfigurasi node rank correlation

■ corr = -1																			
■ corr = +1																			
✗ corr = n/a																			
	date	suhu minimum (°C)	suhu maksimum (°C)	suhu rata-rata (°C)	kelembapan rata-rata (%)	curah hujan (mm)	durasi penirinan matahari (jam)	kecepatan angin maksimum (m/s)	arah angin pada kecepatan maksimum (°)	kecepatan angin rata-rata (m/s)	arah angin tertinggi (°)	station_id	station_id (right)	station_name	region_name	latitude	longitude	region_id	province_id
date	1	-0.005	-0.001	0.004	-0.002	-0.007	0.005	0.004	-0.009	0.003	0.002	0.001	0.001	-0.002	-0	-0.002	-0.002	0.001	0.002
suhu minimum (°C)	-0.005	1	0.371	0.969	-0.324	-0.248	0.083	0.215	-0.156	0.343	0.003	-0.335	-0.335	0.106	0.257	0.566	0.044	-0.029	-0.171
suhu maksimum (°C)	-0.001	0.371	1	0.964	-0.443	-0.285	0.316	0.178	-0.163	0.291	-0.036	-0.33	-0.33	-0.091	0.34	0.311	-0.236	0.174	0.159
suhu rata-rata (°C)	0.004	0.969	0.964	1	-0.588	-0.281	0.298	0.698	-0.287	0.853	0.163	-0.713	-0.713	-0.004	0.894	0.961	-0.632	0.341	0.257
kelembapan rata-rata (%)	-0.002	-0.324	-0.443	-0.588	1	0.486	-0.318	-0.187	0.167	-0.27	0.051	0.13	0.13	-0.174	-0.085	-0.267	-0.038	0.094	0.137
curah hujan (mm)	-0.007	-0.248	-0.285	-0.281	0.486	1	-0.282	-0.093	0.139	-0.134	0.046	0.108	0.108	-0.093	-0.051	-0.193	0.05	0.005	-0.023
durasi penirinan matahari (jam)	0.005	0.083	0.316	0.298	-0.318	-0.282	1	0.081	-0.089	0.084	-0.142	-0.009	-0.009	0.024	0.081	0.087	-0.062	0.028	0.022
kecepatan angin maksimum (m/s)	0.004	0.215	0.178	0.698	-0.187	-0.093	0.081	1	-0.111	0.669	0.191	-0.02	-0.02	0.262	0.216	0.361	-0.231	0.082	0.126
arah angin pada kecepatan ma...	-0.009	-0.156	-0.163	-0.287	0.167	0.139	-0.089	-0.111	1	-0.151	0.043	0.148	0.148	-0.064	-0.13	-0.219	0.071	-0.079	-0.099
kecepatan angin rata-rata (m/s)	0.003	0.343	0.291	0.853	-0.27	-0.134	0.084	0.669	-0.151	1	0.336	-0.226	-0.226	0.15	0.413	0.448	-0.225	0.196	0.18
arah angin tertinggi (°)	0.002	0.003	-0.036	0.163	0.051	0.046	-0.142	0.191	0.043	0.336	1	-0.113	-0.113	0.008	0.115	0.103	-0.128	0.158	0.21
station_id	0.001	-0.335	-0.33	-0.713	0.13	0.108	-0.009	-0.02	0.148	-0.226	-0.113	1	1	0.245	-0.393	-0.381	0.209	-0.358	-0.338
station_id (right)	0.001	-0.335	-0.33	-0.713	0.13	0.108	-0.009	-0.02	0.148	-0.226	-0.113	1	1	0.245	-0.393	-0.381	0.209	-0.358	-0.338
station_name	-0.002	0.106	-0.091	-0.004	-0.174	-0.093	0.024	0.262	-0.064	0.15	0.008	0.245	0.245	1	-0.149	0.07	0.33	-0.386	-0.473
region_name	-0	0.257	0.34	0.894	-0.085	-0.051	0.081	0.216	-0.13	0.413	0.115	-0.393	-0.393	-0.149	1	0.431	-0.375	0.672	0.57
latitude	-0.002	0.566	0.311	0.961	-0.267	-0.193	0.087	0.361	-0.219	0.448	0.103	-0.381	-0.381	0.07	0.431	1	-0.248	0.193	0.151
longitude	-0.002	0.044	-0.236	-0.632	-0.038	0.05	-0.062	-0.231	0.071	-0.225	-0.128	0.209	0.209	0.33	-0.375	-0.248	1	-0.489	-0.773
region_id	0.001	-0.029	0.174	0.341	0.094	0.005	0.028	0.082	-0.079	0.196	0.158	-0.358	-0.358	-0.386	0.672	0.193	-0.489	1	1
province_id	0.002	-0.171	0.159	0.257	0.137	-0.023	0.022	0.126	-0.099	0.18	0.21	-0.338	-0.338	-0.473	0.57	0.151	-0.773	1	1

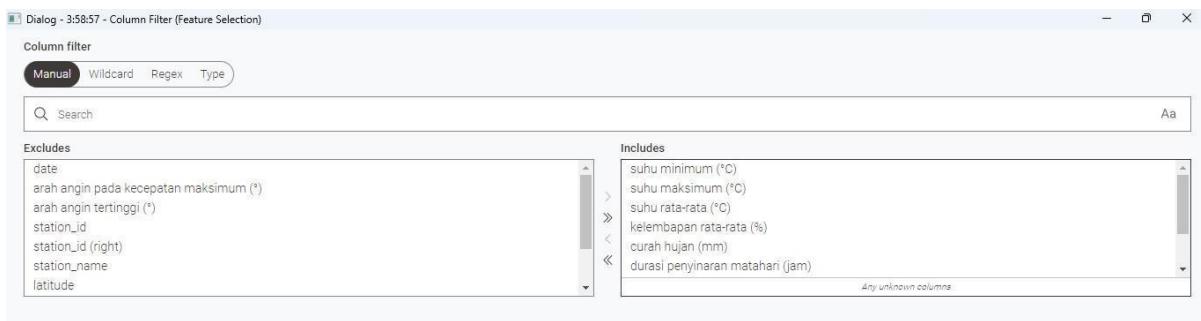
Gambar 4. 54 Hasil *correlation matrix*



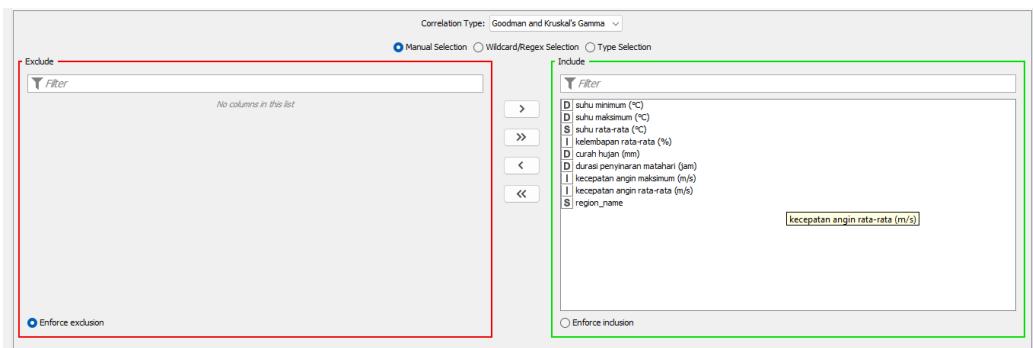
Gambar 4. 55 Rank correlation +Visualisasi Heatmap



Gambar 4. 56 Alur kerja tahap *feature selection* untuk mengurangi fitur



Gambar 4. 57 Konfigurasi untuk pemilihan fitur



Gambar 4. 58 Konfigurasi *rank correlation* yang terpasang setelah *column filter*

	■ corr = -1	■ corr = +1	X corr = n/a	suhu minimum (°C)	suhu maksimum (°C)	suhu rata-rata (°C)	kelembapan rata-rata (%)	curah hujan (mm)	durasi penyinaran matahari (jam)	kecepatan angin maksimum (m/s)	kecepatan angin rata-rata (m/s)	region_name
suhu minimum (°C)	1	0.371	0.969	-0.324	-0.248	0.083	0.215	0.343	0.257			
suhu maksimum (°C)	0.371	1	0.964	-0.443	-0.285	0.316	0.178	0.291	0.34			
suhu rata-rata (°C)	0.969	0.964	1	-0.588	-0.281	0.298	0.698	0.853	0.894			
kelembapan rata-rata (%)	-0.324	-0.443	-0.588	1	0.486	-0.318	-0.187	-0.27	-0.085			
curah hujan (mm)	-0.248	-0.285	-0.281	0.486	1	-0.282	-0.093	-0.134	-0.051			
durasi penyinaran matahari (jam)	0.083	0.316	0.298	-0.318	-0.282	1	0.081	0.084	0.081			
kecepatan angin maksimum (m/s)	0.215	0.178	0.698	-0.187	-0.093	0.081	1	0.669	0.216			
kecepatan angin rata-rata (m/s)	0.343	0.291	0.853	-0.27	-0.134	0.084	0.669	1	0.413			
region_name	0.257	0.34	0.894	-0.085	-0.051	0.081	0.216	0.413	1			

Gambar 4. 59 Hasil korelasi dari fitur yang dipilih



Gambar 4. 60 Rank correlation +Visualisasi Heatmap

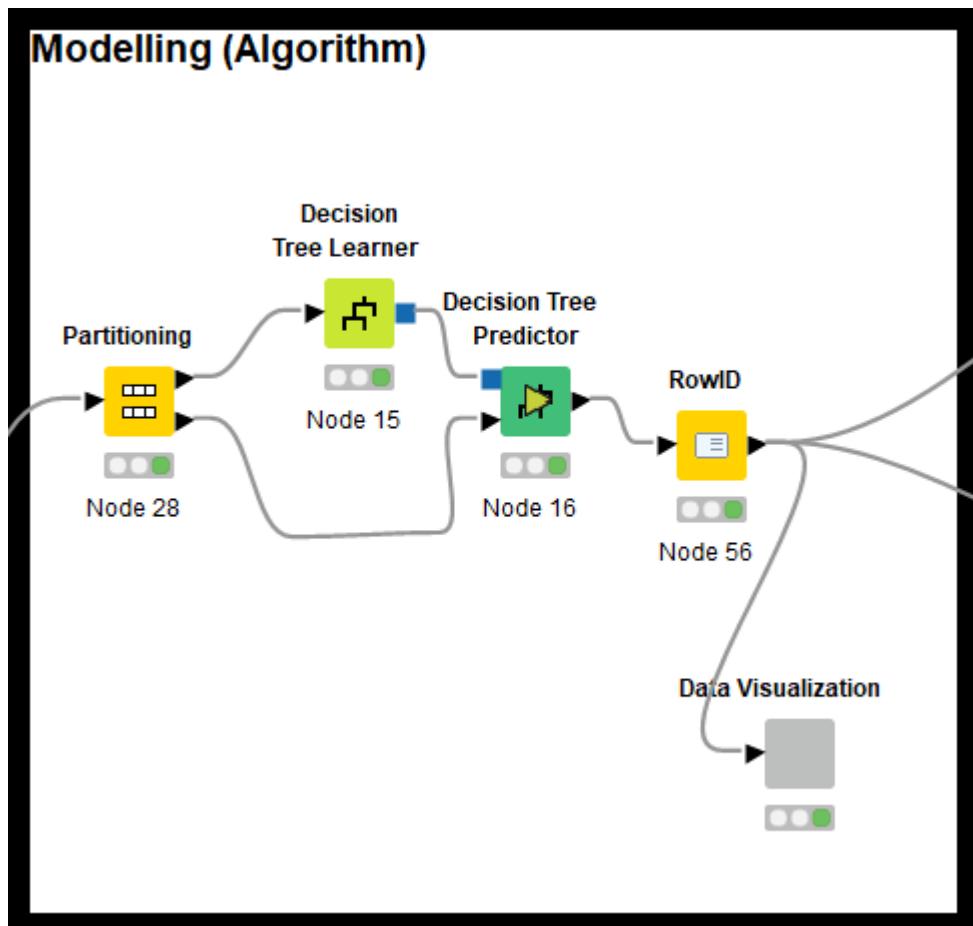
Jadi pemilihan fitur untuk memprediksi suhu rata-rata (°C) terdiri dari 8 kolom. Diantaranya yaitu suhu minimum (°C), suhu maksimum (°C), kelembapan rata-rata (%), curah hujan (mm), durasi penyinaran matahari (jam), kecepatan angin maksimum (m/s), kecepatan angin rata-rata (m/s), dan region_name. Berikut ini adalah alasan pemilihan fitur untuk memprediksi suhu rata-rata yang akan ditampilkan pada Tabel 4. 2 dibawah ini:

Tabel 4. 2 Alasan pemilihan fitur

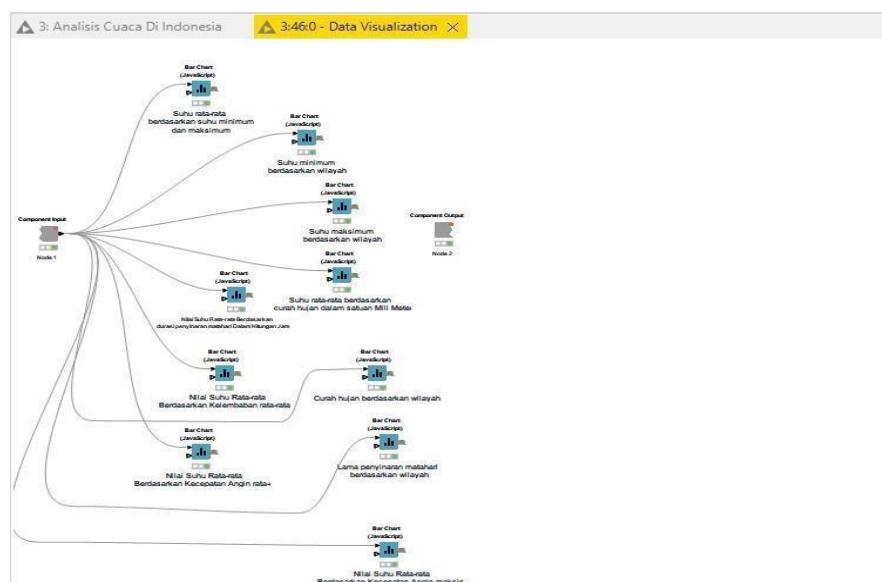
Fitur	Alasan Pemilihan
1. suhu minimum (°C)	Korelasi positif yang sangat kuat
2. suhu maksimum (°C)	Juga sangat kuat berkorelasi positif
3. kelembapan rata-rata (%)	Cenderung negatif, semakin lembab justru suhu menurun
4. curah hujan (mm)	Korelasi negatif ringan hingga kuat
5. durasi penyinaran matahari (jam)	Korelasi positif ringan hingga kuat
6. kecepatan angin maksimum (m/s)	Cenderung positif, semakin cepat angin, suhu pun meningkat
7. kecepatan angin rata-rata (m/s)	Korelasi positif kuat
8. region_name	Korelasi kuat dengan letak suatu wilayah

4.4 Modelling

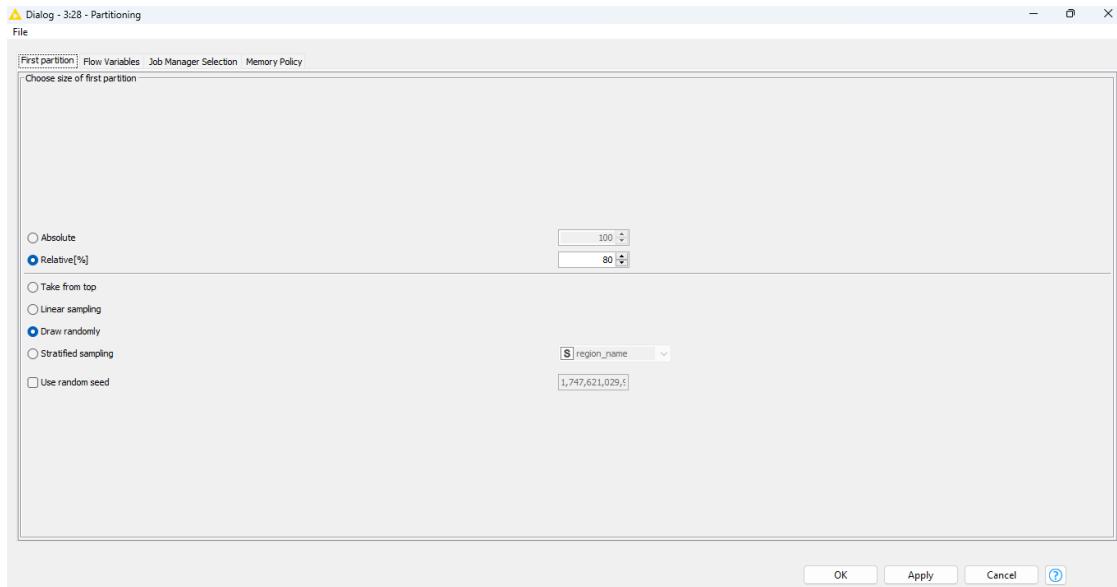
Di tahap ini, proses dimana contoh memakai teknik penjabaran yang dipilih membuat suatu pola info yang bisa memudahkan pihak yang berkepentingan. Pola klasifikasi yang didapatkan teknik data mining ini digunakan buat memprediksi suhu rata-rata dengan predictor yang mempengaruhinya. Adapun tools atau alat bantu di proses pengolahan data mining yaitu Knime Analytics Platform versi 5.2.4. Tahapan ini menampilkan dan menyampaikan informasi terhadap kinerja algoritma. Alur kerja pada tahapan *modelling* yang terdiri dari beberapa node yakni *partitioning*, *decision tree learner*, *decision tree predictor*, *rowid*, dan *component* yang dimana berisi *node* untuk memberikan hasil visualisasi data yang terdiri dari 10 buah *node barchart (Java Script)*. Berikut ini adalah alur kerja pada tahapan *modelling* dibawah ini:



Gambar 4. 61 Alur kerja tahap *modelling*



Gambar 4. 62 Component yang berisi alur kerja untuk visualisasi data



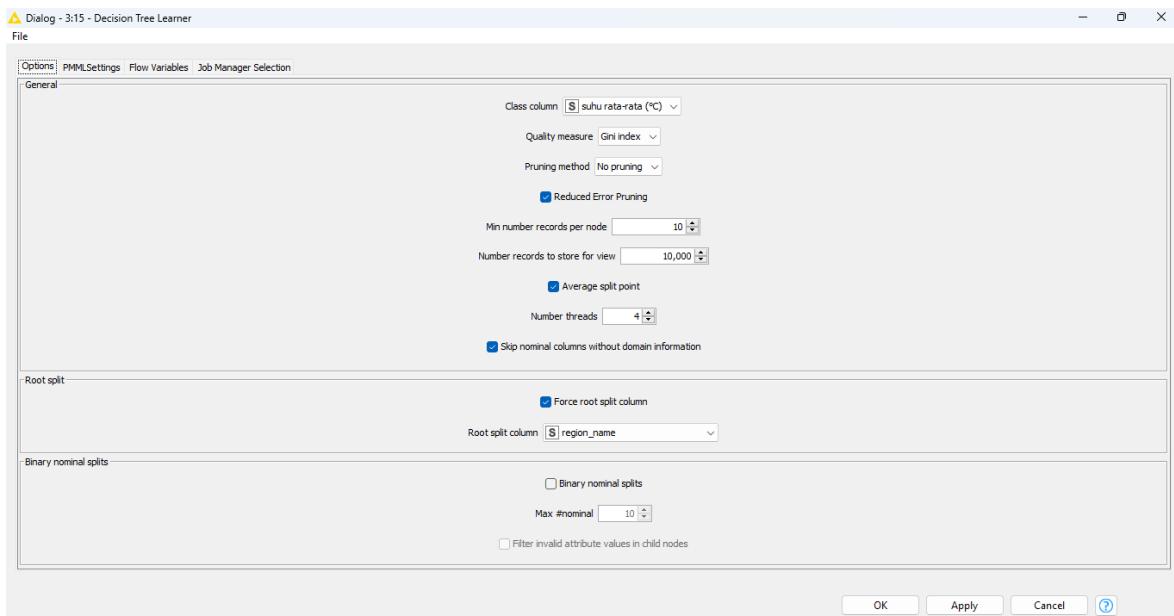
Gambar 4. 63 Konfigurasi *node partitioning* untuk pengujian dan pelatihan data

Row ID	D suhu minimum (°C)	D suhu maksimum (°C)	S suhu rata-rata (°C)	I kelembapan rata-rata (%)	D curah hujan (mm)	D durasi penyinaran matahari (jam)	I kecepatan angin maksimum (m/s)	I kecepatan angin rata-rata (m/s)	S region_name
Row0 25	33.4	> 25	78	0.1	4.8	3	2	Kota Tangerang Selatan	
Row1 25	34.4	> 25	78	0	5.9	4	2	Kota Tangerang Selatan	
Row2 25	33.4	> 25	79	14.6	4.2	4	3	Kota Tangerang Selatan	
Row4 26	32.8	> 25	77	0	1.7	4	3	Kota Tangerang Selatan	
Row5 24	34.8	> 25	76	0	5	3	2	Kota Tangerang Selatan	
Row6 24	29.4	> 25	86	61	0.5	3	0	Kota Tangerang Selatan	
Row7 24	31.4	> 25	84	14	0.7	3	2	Kota Tangerang Selatan	
Row8 24	31.8	> 25	84	21	1.8	3	2	Kota Tangerang Selatan	
Row9 24	32	> 25	87	0	1.9	2	2	Kota Tangerang Selatan	
Row10 24	29.8	> 25	91	5	1	4	3	Kota Tangerang Selatan	
Row11 24	29.8	> 25	87	1	1	3	2	Kota Tangerang Selatan	
Row13 21	31	< 25	88	42.7	1.7	2	1	Kota Tangerang Selatan	
Row14 24	29.4	> 25	91	7	0	4	2	Kota Tangerang Selatan	
Row15 23	32.8	> 25	85	6.9	2.6	3	2	Kota Tangerang Selatan	
Row16 24	30.8	> 25	90	3.2	0.2	5	2	Kota Tangerang Selatan	
Row18 24	33.2	> 25	80	10	4.3	3	2	Kota Tangerang Selatan	
Row19 24	31.2	> 25	89	0	2.2	3	2	Kota Tangerang Selatan	
Row20 24	33.4	> 25	81	9.4	5.8	3	2	Kota Tangerang Selatan	
Row21 25	34	> 25	84	0	7.1	3	2	Kota Tangerang Selatan	
Row22 24	31.4	> 25	84	4	1.8	3	2	Kota Tangerang Selatan	
Row24 24	34	> 25	84	2.2	6.6	3	2	Kota Tangerang Selatan	
Row26 25	33.2	> 25	82	0	6.3	4	2	Kota Tangerang Selatan	
Row29 24	33.6	> 25	82	0.1	3.3	3	2	Kota Tangerang Selatan	
Row30 24	33.8	> 25	83	19	4.1	4	2	Kota Tangerang Selatan	
Row31 25	33.2	> 25	85	0	3.1	3	2	Kota Tangerang Selatan	
Row32 24	33	> 25	82	0.8	4.9	2	0	Kota Tangerang Selatan	
Row33 25	32	> 25	85	1.6	2.7	3	2	Kota Tangerang Selatan	
Row34 25	33	> 25	86	0.5	3	3	2	Kota Tangerang Selatan	
Row35 24	33.8	> 25	80	3	6.5	2	2	Kota Tangerang Selatan	
Row36 25	33.8	> 25	80	0	3.3	3	2	Kota Tangerang Selatan	
Row37 25	35	> 25	75	0	6.2	2	2	Kota Tangerang Selatan	
Row39 25	33.8	> 25	81	0.5	4.8	4	2	Kota Tangerang Selatan	
Row40 25	34	> 25	77	0	4.6	2	2	Kota Tangerang Selatan	
Row41 25	30	> 25	93	0.5	0	2	0	Kota Tangerang Selatan	
Row42 25	33	> 25	83	7.4	1.9	3	2	Kota Tangerang Selatan	
Row43 23	34.8	> 25	81	2	5.9	3	2	Kota Tangerang Selatan	
Row45 24	33.6	> 25	86	7.1	4.2	2	2	Kota Tangerang Selatan	
Row46 24	30.4	> 25	90	1.8	1.4	2	1	Kota Tangerang Selatan	
Row48 24	33.4	> 25	88	3	4.1	4	2	Kota Tangerang Selatan	

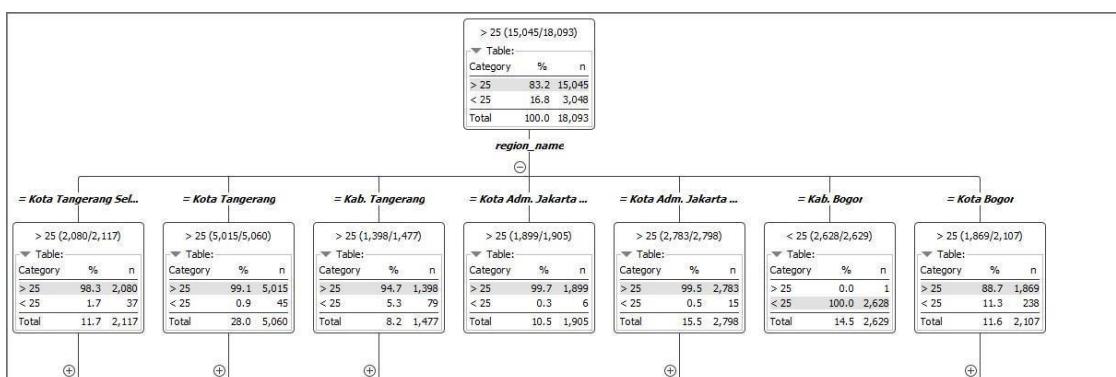
Gambar 4. 64 Tampilan tabel data setelah pengujian dan pelatihan data

Penjelasan Gambar 4. 64 *node "Partitioning"*, di mana 80% pelatihan dicadangkan untuk pengujian 20%. 80% dari set data yang dicadangkan untuk pelatihan diberikan kepada *node Decision Tree Learner*. Pada pengujung algoritma *node Decision Tree Learner* dihubungkan ke *node Predictor* masukan pertama yang memiliki dua masukan

dan masukan lain dari *Decision Tree Predictor* mengambil data dari simpul *Partitioning*. Dari data terlatih (*data training*) tersebut, *Decision Tree Predictor* sekarang diperlukan untuk memprediksi data uji. Sebagaimana akan dipaparkan pada Gambar 4.84 hingga Gambar 4.65 berikut ini:



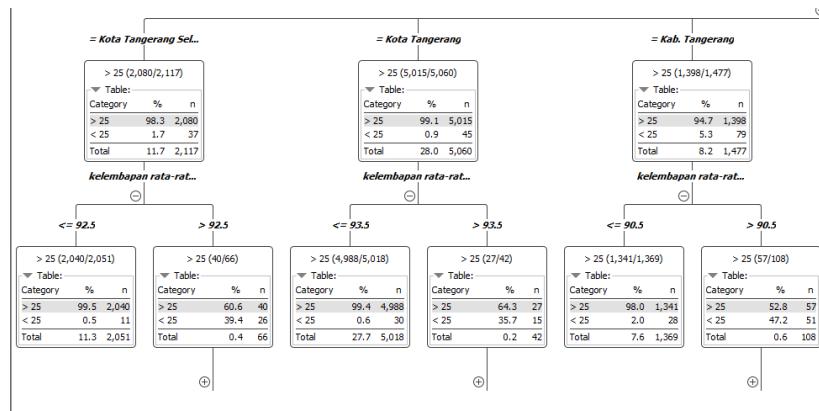
Gambar 4. 65 Konfigurasi node decision tree learner tab options



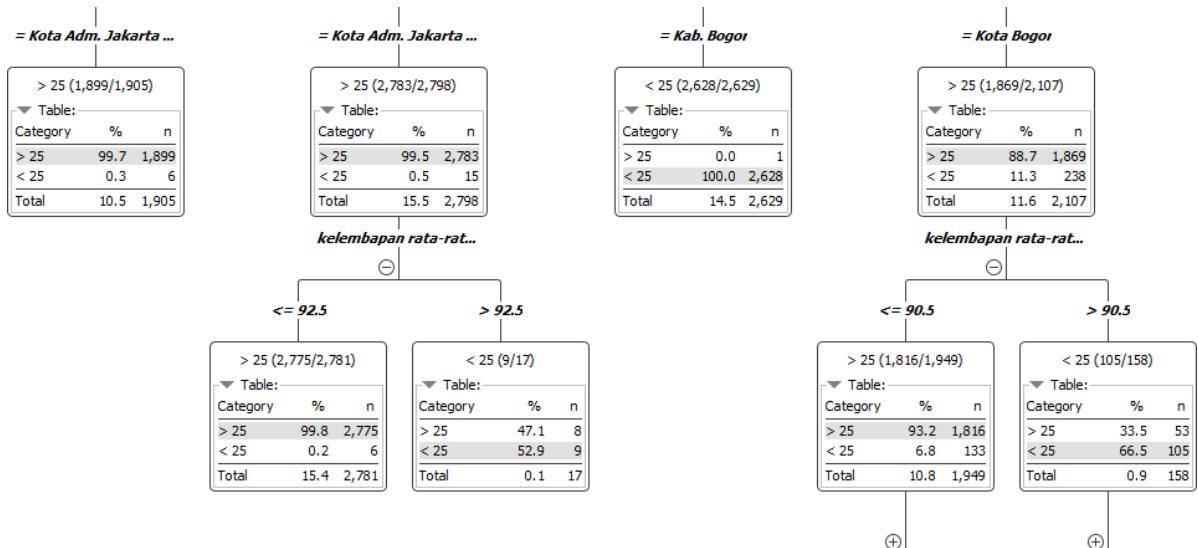
Gambar 4. 66 Tampilan *decision tree* berdasarkan wilayah

Penjelasan pada Gambar 4. 66 adalah bahwasannya sebagian besar wilayah yang dianalisis yakni suhu rata-rata yang mencapai lebih dari 25°C yakni 83,2 %. Sementara itu, yang suhu rata-ratanya kurang dari 25°C hanya sebagian kecil saja yakni 16,8%.

Selanjutnya pemecahan berdasarkan wilayah (*region_name*) yang dimana kota Adm. Jakarta Utara wilayah yang tergolong panas yakni 99,7% memiliki suhu rata-rata yang lebih dari 25°C dibandingkan wilayah lainnya. Sementara itu, dimana Kabupaten Bogor wilayah yang tergolong dingin yakni 100% memiliki suhu rata-rata yang kurang dari 25°C dibandingkan wilayah lainnya.



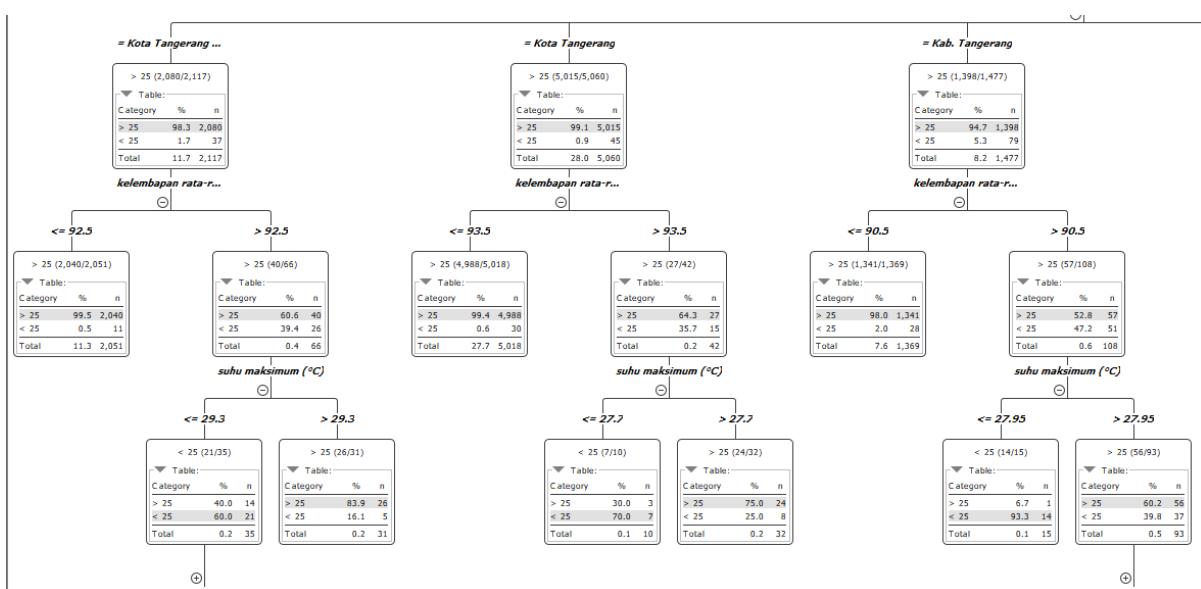
Gambar 4.67 Tampilan *decision tree* berupa pemecahan suhu rata-rata suatu wilayah berdasarkan kelembaban rata-rata



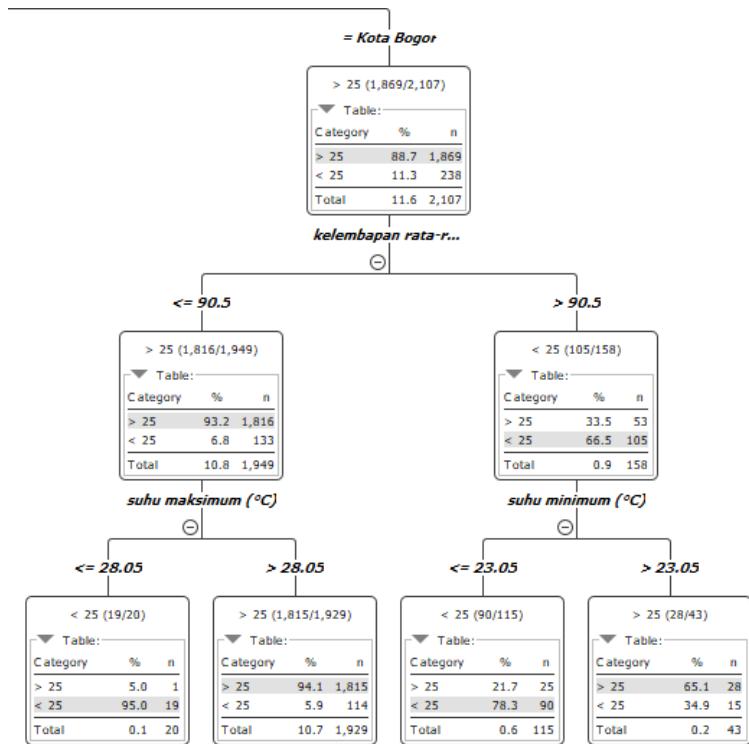
Gambar 4. 68 Tampilan *decision tree* berupa pemecahan suhu rata-rata suatu wilayah berdasarkan kelembaban rata-rata

Penjelasan dari kedua gambar yakni Gambar 4. 77 dan Gambar 4. 78 merupakan pemecahan suhu rata-rata suatu wilayah berdasarkan kelembaban rata-rata yang dimana

kota atau wilayah yang tergolong lembab yaitu terletak di kota Tangerang yakni 93,5% dibandingkan kota atau wilayah lainnya yang terdapat pada Gambar 4. 98. Kelembaban rata-rata di wilayah tersebut yang lebih dari 93,5% terdapat 64,3% suhu rata-rata yang melebihi 25°C. Sementara itu, kelembaban rata-rata yang kurang dari atau sama dengan 93,5% terdapat 99,4% suhu rata-rata yang melebihi 25°C yang berarti sebagian besar wilayah cenderung tidak terlalu lembab dikarenakan suhu cenderung tinggi.

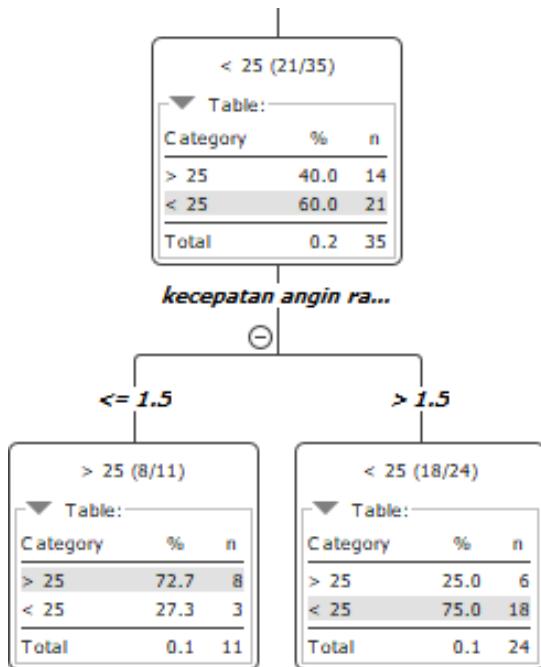


Gambar 4. 69 Tampilan *decision tree* berupa pemecahan suhu rata-rata di suatu wilayah berdasarkan suhu maksimum

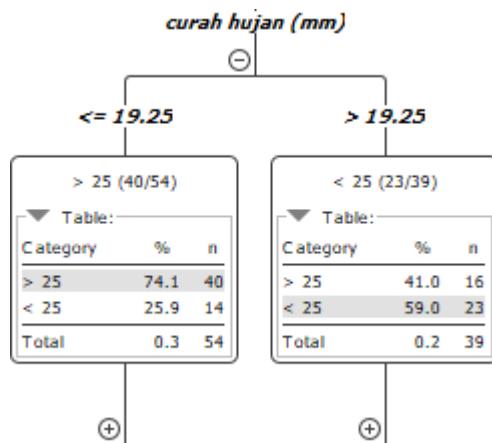


Gambar 4. 70 Tampilan *decision tree* berupa pemecahan suhu rata-rata di suatu wilayah berdasarkan suhu maksimum dan minimum

Penjelasan Gambar 4. 70 merupakan perbandingan pemecahan suhu rata-rata berdasarkan antara suhu maksimum dan suhu minimum. Jika suhu maksimum bertambah maka suhu rata rata pun meningkat begitu pula sebaliknya. Seperti yang dipaparkan pada Gambar 4.101 terdapat suhu maksimum lebih dari 28,05°C sudah dipastikan sebagian besar suhu rata-ratanya lebih dari 25°C yakni mencapai 94,1% dibandingkan suhu rata-rata yang kurang dari 25°C. Lalu terdapat suhu maksimum yang kurang dari 28,05°C sebagian besar suhu rata-ratanya kurang dari 25°C yakni mencapai 95% dibandingkan suhu rata-rata yang lebih dari 25°C. Selanjutnya berdasarkan suhu minimum jika bertambah maka suhu rata rata pun menurun begitu pula sebaliknya. Terdapat suhu minimum kurang dari 23,05°C sudah dipastikan sebagian besar suhu rata-ratanya kurang dari 25°C yakni mencapai 78,3% dibandingkan suhu rata-rata yang lebih dari 25°C. Lalu terdapat suhu minimum yang lebih dari 23,05°C hampir sebagian besar suhu rata-ratanya lebih dari 25°C yakni mencapai 65,1% dibandingkan suhu rata-rata yang kurang dari 25°C.

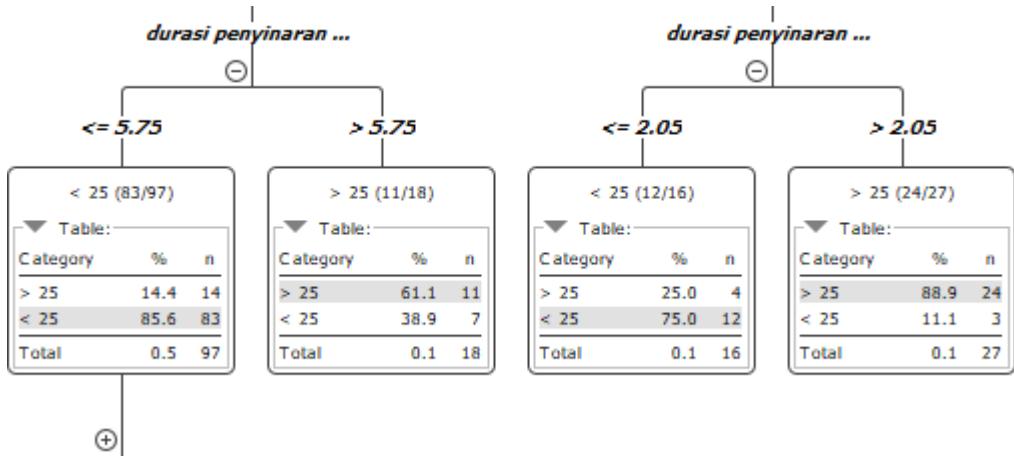


Gambar 4. 71 Tampilan *decision tree* berupa pemecahan suhu rata-rata berdasarkan kecepatan angin rata-rata



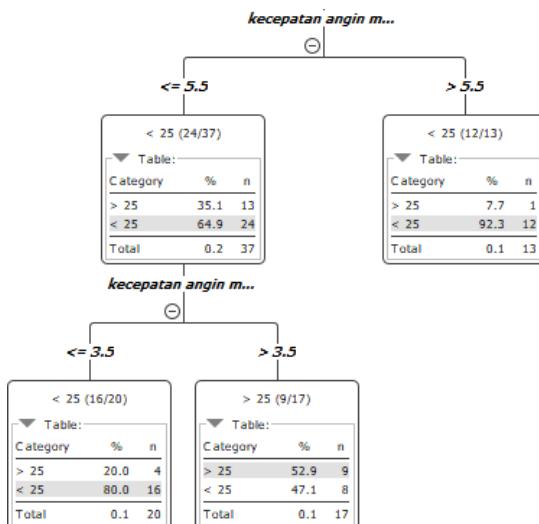
Gambar 4. 72 Tampilan *decision tree* berupa pemecahan suhu rata-rata berdasarkan curah hujan dalam satuan milimeter

Penjelasan pada Gambar 4. 72 adalah bahwasannya curah hujan jika semakin lebat, maka suhu rata-rata pun menurun. Sementara itu, jika curah hujan tidak terlalu lebat, maka suhu rata-rata pun meningkat. Seperti yang dipaparkan pada Gambar 4. 82 yaitu hampir sebagian besar curah hujan yang berukuran lebih dari 19,25 mm yaitu suhu rata-rata yang kurang dari 25°C. Sedangkan sebagian besar curah hujan yang berukuran kurang dari 19,25 mm yaitu suhu rata-rata yang lebih dari 25°C.

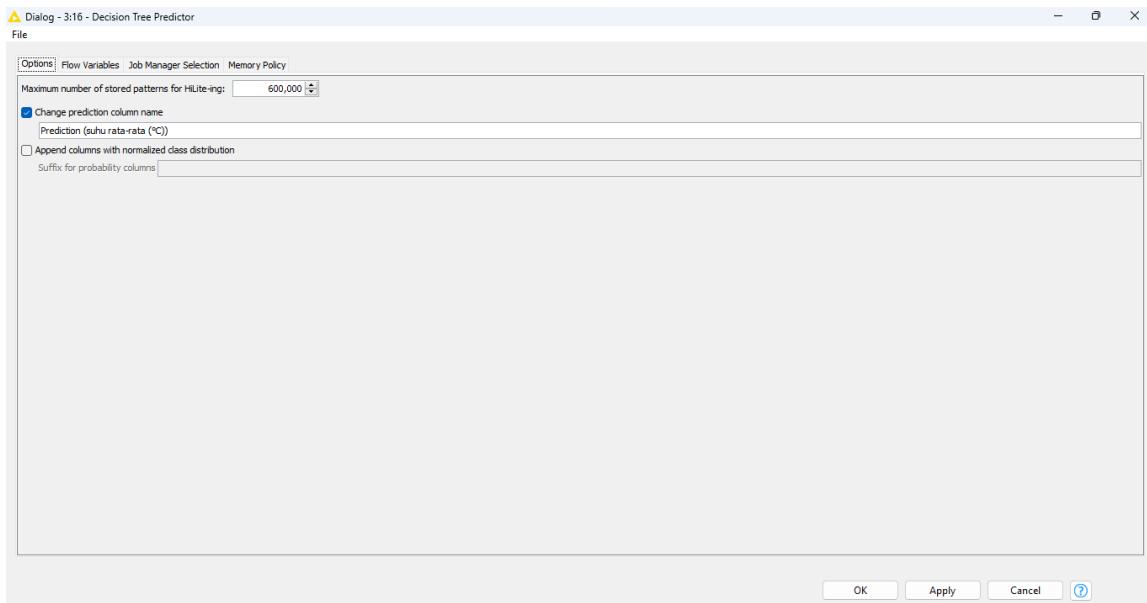


Gambar 4. 73 Tampilan *decision tree* berupa pemecahan suhu rata-rata beserta suhu minimum berdasarkan durasi penyinaran matahari

Penjelasan Gambar 4. 73 adalah bahwasannya jika durasi sinar matahari semakin bertambah, maka suhu rata-rata pun meningkat. Sementara itu, jika durasi sinar matahari semakin berkurang, maka suhu rata-rata pun menurun. Seperti yang dipaparkan pada Gambar 4. 83 yaitu hampir sebagian besar suhu rata-rata lebih dari 25°C jika lama penyinaran matahari berdurasi lebih tinggi. Sedangkan jika lama penyinaran matahari yang berdurasi tidak terlalu tinggi, maka hampir hingga sebagian besar suhu rata-rata kurang dari 25°C.



Gambar 4. 74 tampilan *decision tree* berupa pemecahan suhu rata-rata berdasarkan kecepatan angin maksimum

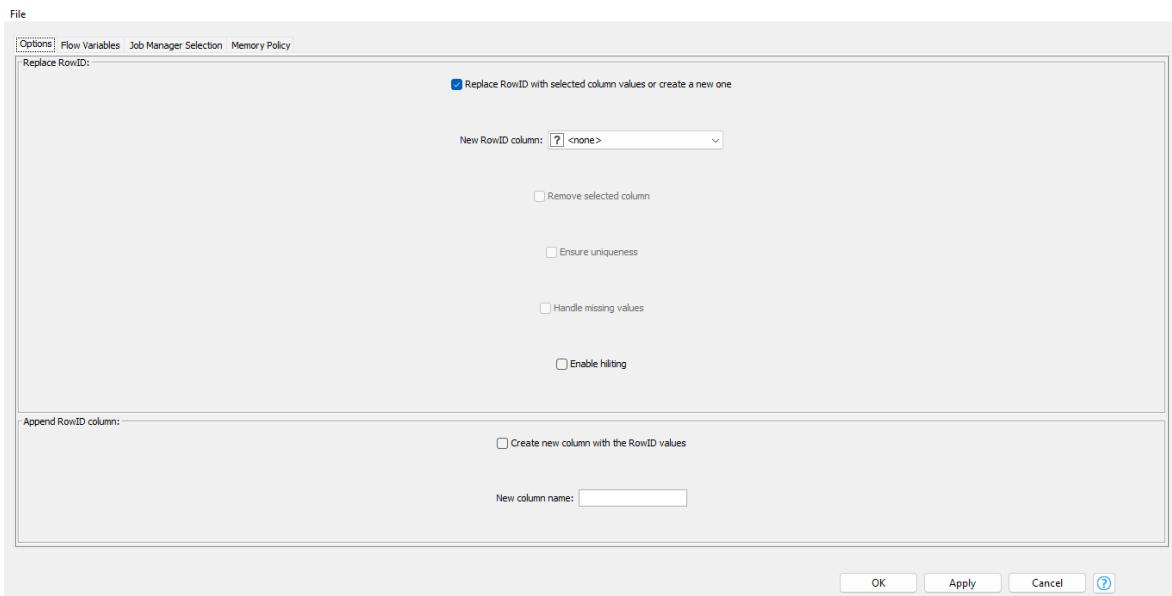


Gambar 4. 75 Konfigurasi *decision tree predictor*

Row ID	D suhu minimum (°C)	D suhu maksimum (°C)	S suhu rata-rata (°C)	I kelembapan rata-rata (%)	D curah hujan (mm)	D durasi peninjoran matahari (jam)	I kecepatan angin maksimum (km/jam)	I kecepatan angin rata-rata (km/jam)	S region_name	S Predict...
Row3 26	34.4	> 25	72	0	8	5	4	Kota Tangerang Selatan > 25		
Row12 23	29.4	> 25	96	8	2.5	3	2	Kota Tangerang Selatan > 25		
Row16 23	30	> 25	89	20.3	1	3	2	Kota Tangerang Selatan > 25		
Row23 24	31.2	> 25	86	5.2	1	3	2	Kota Tangerang Selatan > 25		
Row24 24	33.6	> 25	78	3.2	5.8	3	2	Kota Tangerang Selatan > 25		
Row27 23	33	> 25	80	9.6	6.4	3	2	Kota Tangerang Selatan > 25		
Row28 24	31	> 25	98	15.5	0	1	0	Kota Tangerang Selatan > 25		
Row38 24	34.2	> 25	78	0.5	5.7	3	2	Kota Tangerang Selatan > 25		
Row44 24	32.4	> 25	87	1.3	4.4	3	2	Kota Tangerang Selatan > 25		
Row47 24	32	> 25	88	59.5	0.7	3	2	Kota Tangerang Selatan > 25		
Row49 24	34	> 25	83	18.5	7.9	2	1	Kota Tangerang Selatan > 25		
Row57 25	35	> 25	83	19.5	5.1	5	2	Kota Tangerang Selatan > 25		
Row58 24	35	> 25	78	2.3	7.7	3	2	Kota Tangerang Selatan > 25		
Row63 25	34.6	> 25	79	5	7	3	2	Kota Tangerang Selatan > 25		
Row64 25	34	> 25	90	0	3.7	3	1	Kota Tangerang Selatan > 25		
Row73 25	32.8	> 25	84	2.2	3.9	3	2	Kota Tangerang Selatan > 25		
Row74 24	34	> 25	81	1.5	6.9	2	1	Kota Tangerang Selatan > 25		
Row90 25	35.2	> 25	81	1.7	4.6	2	2	Kota Tangerang Selatan > 25		
Row92 24	34.6	> 25	74	0	4	3	1	Kota Tangerang Selatan > 25		
Row94 25	35.8	> 25	80	17.7	7.3	4	2	Kota Tangerang Selatan > 25		
Row102 24	34.4	> 25	73	0	6.5	3	2	Kota Tangerang Selatan > 25		
Row103 24	33.8	> 25	78	0	5.8	2	1	Kota Tangerang Selatan > 25		
Row104 24	34.6	> 25	71	0	8	5	4	Kota Tangerang Selatan > 25		
Row108 26	34.2	> 25	78	0	1	3	2	Kota Tangerang Selatan > 25		
Row109 25	35.4	> 25	76	0	6.2	2	1	Kota Tangerang Selatan > 25		
Row118 27	33.4	> 25	78	0	2.4	4	2	Kota Tangerang Selatan > 25		
Row119 25	33.4	> 25	77	0	4.1	2	2	Kota Tangerang Selatan > 25		
Row129 25	31.4	> 25	92	14	0.6	3	0	Kota Tangerang Selatan > 25		
Row130 26	34.8	> 25	77	0	5.1	3	2	Kota Tangerang Selatan > 25		
Row141 25	33.6	> 25	78	0	5.3	3	2	Kota Tangerang Selatan > 25		
Row142 25	34.2	> 25	79	0	7.3	3	2	Kota Tangerang Selatan > 25		
Row143 25	34	> 25	80	0	4.4	2	2	Kota Tangerang Selatan > 25		
Row155 25	33.6	> 25	82	0	5.2	3	2	Kota Tangerang Selatan > 25		
Row163 25	32.2	> 25	77	3.8	4.7	4	2	Kota Tangerang Selatan > 25		
Row166 24	34	> 25	74	0	4.1	2	2	Kota Tangerang Selatan > 25		
Row167 25	32.6	> 25	84	2.4	2.9	3	2	Kota Tangerang Selatan > 25		
Row169 25	32	> 25	86	0.9	2.3	2	1	Kota Tangerang Selatan > 25		
Row170 23	33.2	> 25	81	47.5	7	3	1	Kota Tangerang Selatan > 25		
Row171 25	31.2	> 25	85	0	0.6	0	0	Kota Tangerang Selatan > 25		

Gambar 4. 76 Tampilan tabel data setelah dikonfigurasi menggunakan *predictor*

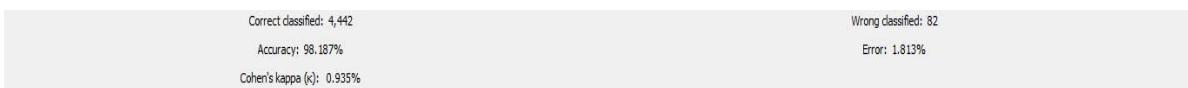
Penjelasan Gambar 4. 75 dan Gambar 4. 76 yaitu konfigurasi *node decision tree predictor* yang dimana berfungsi untuk memprediksi model *decision tree* yang sebelumnya sudah dilatih ke data baru. *Node* ini digunakan pada saat ingin menguji model pada pengujian data. Dan juga untuk menampilkan kolom baru berupa *output* yang dimana *output* tersebut berisikan hasil prediksi.



Gambar 4. 77 Konfigurasi untuk menetapkan ulang Id baris

4.5 Evaluation

Pada tahap evaluation dilakukan evaluasi terhadap model *decision tree* yang dihasilkan dari 4524 data, nilai *true positive* untuk prediksi tersebut adalah 4442 data yang terbukti sempurna serta sesuai dengan kenyataan. Sementara itu terdapat 82 data yang tidak sempurna (*False Positive*)[26]. Hasil akurasi dari model menggunakan aplikasi Knime diperlihatkan pada Gambar 4. 78 dan Gambar 4. 79 sebagai berikut:



Gambar 4. 78 Hasil akurasi model *decision tree*

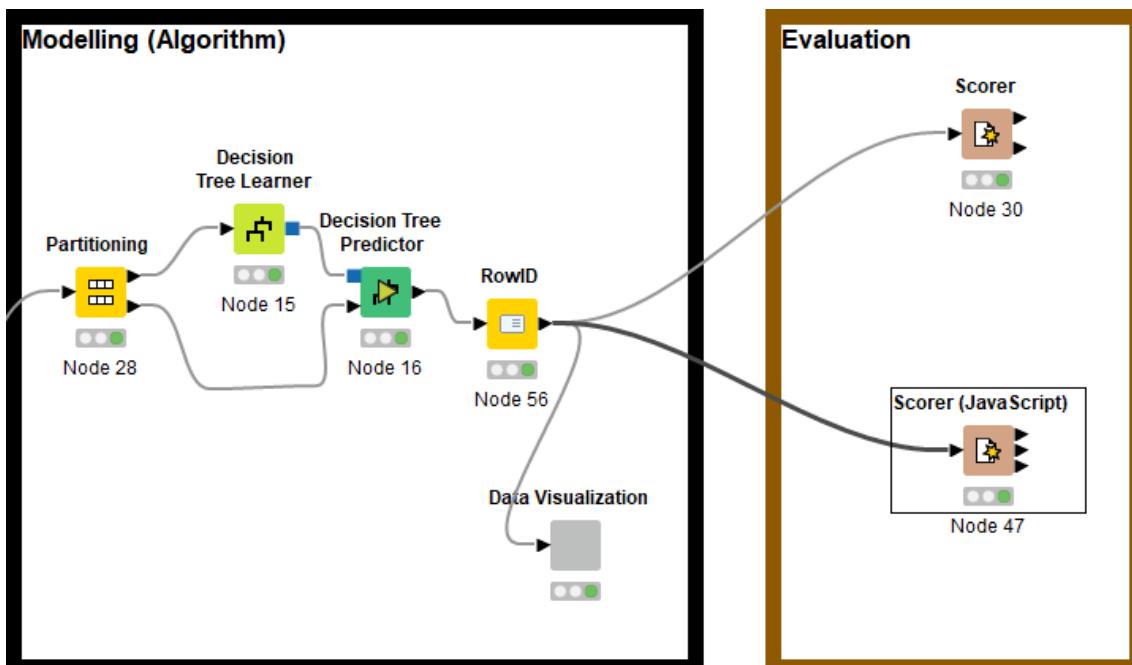
Scorer View

Confusion Matrix

		< 25 (Predicted)	> 25 (Predicted)	
< 25 (Actual)		716	56	92.75%
> 25 (Actual)		26	3726	99.31%
		96.50%	98.52%	
Overall Statistics	Overall Error	Cohen's kappa (κ)	Correctly Classified	Incorrectly Classified
98.19%	1.81%	0.935	4442	82

Gambar 4. 79 Hasil akurasi *confusion matrix*

Hasil evaluasi akurasi menggunakan *confusion matrix* ini yaitu untuk melihat hasil akurasi dari model algoritma yang kita pilih. Terdapat beberapa *node* dari tahapan *evaluation* yaitu *node scorer* dan *scorer (Java script)* untuk membulatkan hasil akurasi yang dimana menggunakan *node scorer* mencapai 98,187% dibulatkan menjadi 98,19%. Alur kerja pada tahapan *evaluation* dipasangkan setelah alur kerja tahap *modelling*. Sebagaimana yang akan dipaparkan pada Gambar 4. 80 dibawah ini:



Gambar 4. 80 Alur kerja tahapan *evaluation*

Akurasi dari model ini juga dapat dihitung menggunakan rumus dibawah ini:

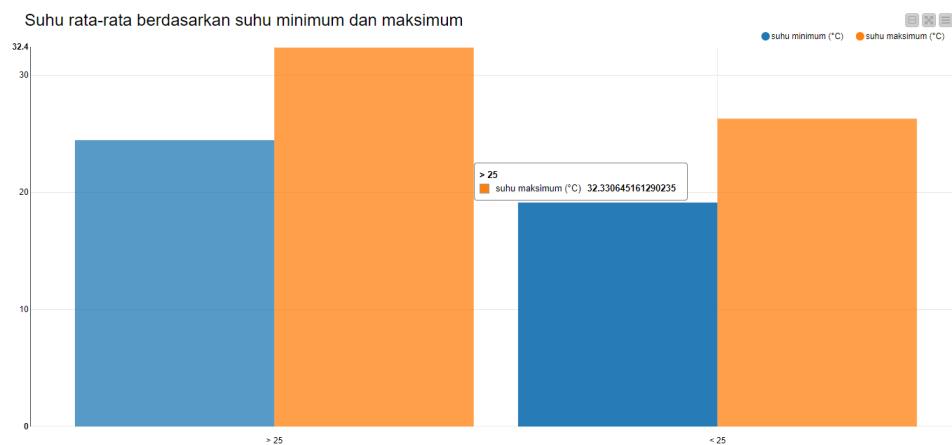
$$Akurasi = \left(\frac{Jumlah Tepat}{Jumlah Prediksi} \right) * 100\%$$

$$Akurasi = \left(\frac{4442}{4524} \right) * 100\%$$

$$Akurasi = 98.187\%$$

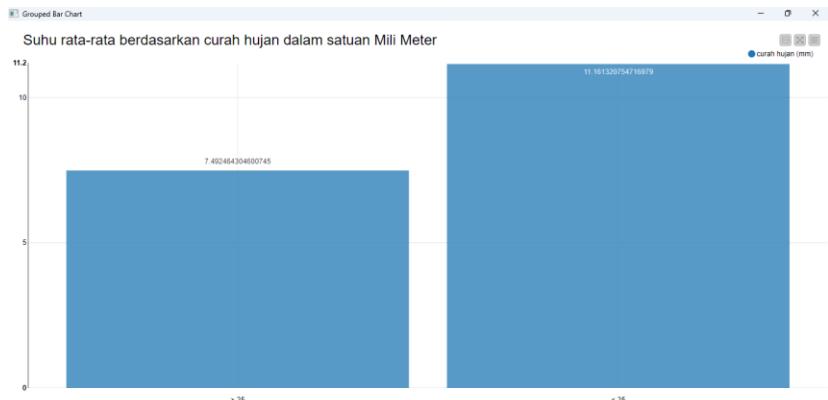
Dengan nilai error = 1,813%

Hasil model prediksi suhu rata-rata berdasarkan suhu minimum dan suhu maksimum diperlihatkan pada Gambar 4. 100 dibawah ini:



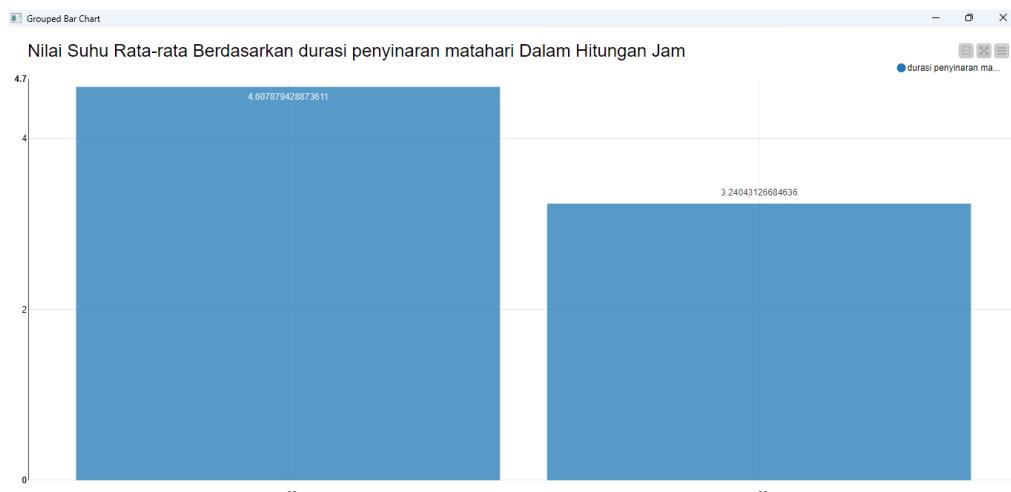
Gambar 4. 81 Grafik suhu rata-rata berdasarkan suhu minimum dan maksimum

Dari Gambar 4. 81 diperlihatkan grafik perbandingan suhu rata-rata berdasarkan antara suhu maksimum dan suhu minimum. Jika suhu rata-rata lebih dari 25°C maka suhu meningkat baik suhu minimum maupun maksimum. Sebaliknya jika suhu rata-rata kurang dari 25°C maka suhu menurun baik suhu maksimum maupun minimum.



Gambar 4. 82 Grafik suhu rata-rata berdasarkan curah hujan dalam satuan Mili Meter

Penjelasan Gambar 4. 82 adalah bahwa jika curah hujan tinggi, maka suhu rata-rata yang kurang dari 25°C meningkat. Dan jika curah hujan menurun, maka suhu rata-rata yang lebih dari 25°C menurun. Seperti yang telah dipaparkan pada Gambar 4. 82 sebelumnya.



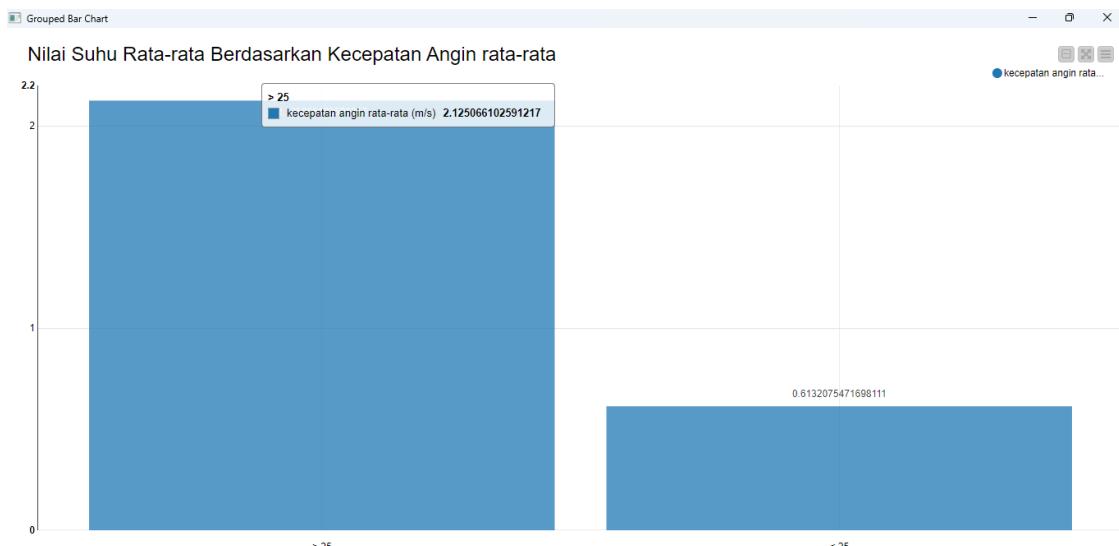
Gambar 4. 83 Grafik Suhu Rata-rata Berdasarkan durasi peninjauan matahari Dalam Hitungan Jam

Penjelasan Gambar 4. 93 merupakan hasil visualisasi data atau grafik suhu rata-rata berdasarkan durasi peninjauan matahari dalam hitungan jam. Jika durasi sinar matahari semakin lama maka suhu rata-rata pun meningkat. Nilai suhu rata-rata tertinggi berdasarkan durasi sinar matahari yaitu lebih dari 25°C yang dimana rata-rata durasi peninjauan matahari adalah 4,6 jam.



Gambar 4. 84 Grafik Suhu Rata-rata Berdasarkan Kelembaban rata-rata

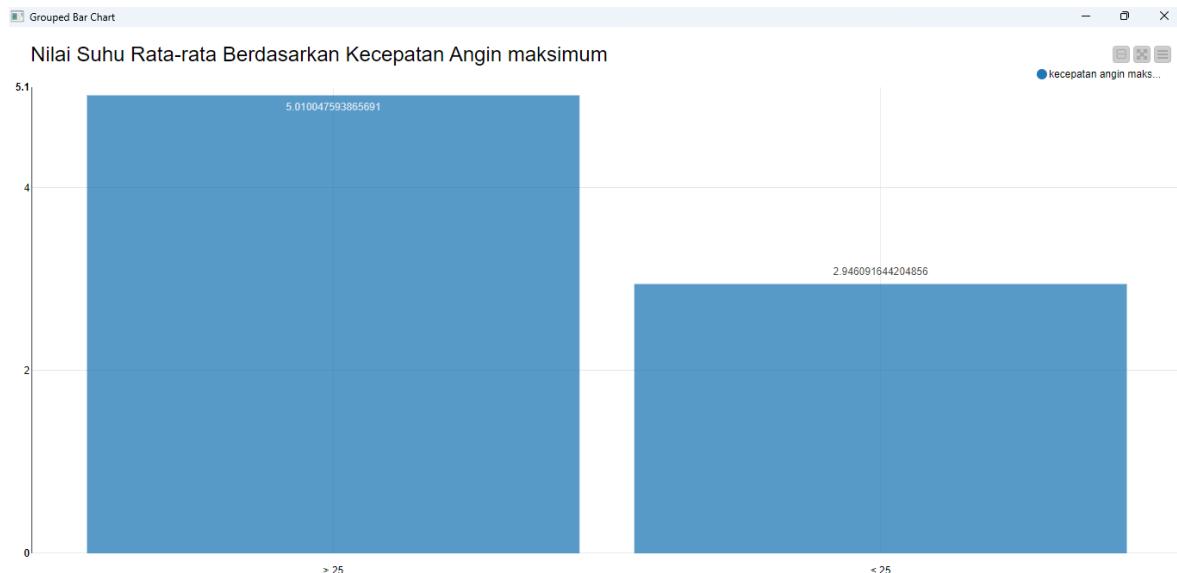
Penjelasan Gambar 4. 84 merupakan hasil visualisasi data atau grafik suhu rata-rata berdasarkan kelembaban rata-rata. Biasanya jika suhu menurun maka akan bertambah lembab, sebaliknya jika suhu naik maka kelembaban pun menurun. Maka hasilnya yaitu suhu rata-rata yang kurang dari 25°C menjadi kelembaban rata-rata yang tinggi yaitu mencapai 86% dibandingkan dengan suhu rata-rata yang lebih dari 25°C.



Gambar 4. 85 Grafik suhu rata-rata berdasarkan kecepatan angin rata-rata

Penjelasan Gambar 4. 85 merupakan hasil visualisasi data atau grafik suhu rata-rata berdasarkan kecepatan angin rata-rata. Jika kecepatan angin meningkat maka suhu pun

sedikit meningkat. Nilai suhu rata-rata tertinggi berdasarkan kecepatan angin rata-rata yaitu lebih dari 25°C yang dimana rata-rata kecepatan angin adalah kurang lebih 2,1 m/s.



Gambar 4. 86 Grafik suhu rata-rata berdasarkan kecepatan angin maksimum

Penjelasan Gambar 4. 86 merupakan hasil visualisasi data atau grafik suhu rata-rata berdasarkan kecepatan angin maksimum. Jika kecepatan angin meningkat maka suhu pun meningkat dan tidak terlalu signifikan. Nilai suhu rata-rata tertinggi berdasarkan kecepatan angin maksimum yaitu lebih dari 25°C yang dimana rata-rata kecepatan angin maksimum adalah kurang lebih 5 m/s.

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil penelitian yang dilakukan terhadap prediksi cuaca di Jakarta, Bogor, dan Tangerang menggunakan algoritma *Decision Tree* pada Knime *Analytics Platform*, yang diperoleh beberapa kesimpulan yang secara langsung menjawab rumusan masalah penelitian dibawah ini:

1. Prediksi cuaca di Jakarta, Bogor, dan Tangerang dengan algoritma Decision Tree menggunakan KNIME analytics platform dilakukan menggunakan metode CRISP-DM meliputi tahapan *business understanding*, *data understanding*, *data preparation*, *modelling*, dan *evaluation*. Dataset cuaca pada pemodelan bersumber dari BMKG pada periode 2010 sampai dengan 2020 dengan pembagian 80% data training dan 20% data testing. Algoritma decision tree pada aplikasi KNIME diterapkan dengan *node partitioning*, *node decision tree learner*, *node decision tree predictor*, dan *component* yang berisi 10 *node data visualization* yaitu *node bar chart (Java Script)*.
2. Model prediksi cuaca menggunakan algoritma decision tree berhasil diterapkan pada data testing dengan akurasi model mencapai 98,187%, dengan nilai *cohen's kappa* mencapai 0,935 artinya model telah memiliki nilai kesepakatan prediksi yang hampir sempurna atau sangat tinggi. Dibandingkan dengan wilayah lainnya, hasil prediksi model cuaca menyatakan menyatakan bahwa wilayah Jakarta Utara 99,7% memiliki suhu panas dengan rata-rata diatas 25°C, sedangkan prediksi cuaca wilayah Kabupaten Bogor memiliki suhu tidak panas (dingin) sebesar 100% dengan suhu dibawah rata-rata kurang dari 25°C.

5.2 Saran

Berdasarkan hasil penelitian dan kesimpulan yang diperoleh, berikut ini adalah beberapa saran yang dapat diajukan adalah sebagai berikut:

1. Peneliti selanjutnya dapat mengimplementasikan menggunakan algoritma lain pada metode *classification* seperti *Naive Bayes*, *Logistic Regression*, *K-nearest Neighbor (KNN)*, *Random Forest*, dan *Support Vector Machine (SVM)*.
2. Dapat mengimplementasikan ke tahapan pemrograman seperti *seaborn*, *numpy*, *pandas*, dan lain-lain.
3. Karena peneliti ini memprediksi hanya di wilayah Jakarta, Bogor, dan Tangerang, sebaiknya ditambahkan 2 wilayah yaitu Depok dan Bekasi agar penelitian nya lengkap menjadi prediksi cuaca harian di wilayah Jabodetabek.
4. Pada hasil penelitian ini, penulis berharap agar penelitian ini bisa dikembangkan oleh peneliti lain dimasa yang akan tiba.

DAFTAR REFERENSI

- [1] J. Junaidi and K. Ramadhani, “Efektivitas Internet of Things (Iot) Pada Sektor Pertanian,” *J. Tek.*, vol. 4, no. 1, p. 12, 2024, doi: 10.54314/teknisi.v4i1.1793.
- [2] Y. Armi, *Berita Cuaca dalam Hubungannya dengan Keselamatan Pelayaran (Studi Kasus Karamnya MV.Xing Shun 01)*. 2023. [Online]. Available: http://repository.pip-semarang.ac.id/5136/2/551811126610N_SKRIPSI_OPEN_ACCESS.pdf
- [3] M. Y. R. Rangkuti, M. V. Alfansyuri, and W. Gunawan, “Penerapan Algoritma K-Nearest Neighbor (Knn) Dalam Memprediksi Dan Menghitung Tingkat Akurasi Data Cuaca Di Indonesia,” *Hexag. J. Tek. dan Sains*, vol. 2, no. 2, pp. 11–16, 2021, doi: 10.36761/hexagon.v2i2.1082.
- [4] M. Maulita and N. Nurdin, “Pendekatan Data Mining Untuk Analisa Curah Hujan Menggunakan Metode Regresi Linear Berganda (Studi Kasus: Kabupaten Aceh Utara),” *IDEALIS Indones. J. Inf. Syst.*, vol. 6, no. 2, pp. 99–106, 2023, doi: 10.36080/idealis.v6i2.3034.
- [5] M. Fauzi *et al.*, “Implementasi Machine Learning Untuk Memprediksi Cuaca Menggunakan Support Vector Machine,” *J. Ilm. Komputasi*, vol. 23, no. 1, pp. 45–50, 2024, doi: 10.32409/jikstik.23.1.3499.
- [6] R. Risanti, “Analisis Model Prediksi Cuaca Menggunakan Support Vector Machine, Gradient Boosting, Random Forest, Dan Decision Tree,” vol. XII, pp. 119–128, 2024, doi: 10.21009/03.1201.fa18.
- [7] K. Tri, W. Ramadhani, D. Pembimbing, D. T. Fisika, and F. T. Industri, “Implementasi Machine Learning Dalam Prediksi Cuaca Maritim Dengan Menggunakan Algoritma Artificial Neural Network Lavenberg-,” 2019.
- [8] S. Shalsabilla, P. Rachmawati, K. Vidya Prakusa, and S. Rihastuti, “Penerapan Data Mining dengan Metode Decision Tree untuk Prediksi Cuaca di Kota Seattle menggunakan Aplikasi Weka,” *Semin. Nas. Amikom Surakarta*, no. November, pp. 93–100, 2023.
- [9] J. Jurnal Mitra Teknik Sipil, “Daftar Isi,” *JMTS J. Mitra Tek. Sipil*, vol. 4, no. 2, 2021, doi: 10.24912/jmts.v0i0.12637.
- [10] N. Made, R. Amanda, M. A. Pramudita, and J. E. Silalahi, “Penggunaan Media Sosial Instagram dalam Komunikasi Bencana Kesiapsiagaan Erupsi Gunung Merapi,” vol. 4, no. 2, pp. 151–165, 2024.

- [11] S. P. Dewi, N. Nurwati, and E. Rahayu, “Penerapan Data Mining Untuk Prediksi Penjualan Produk Terlaris Menggunakan Metode K-Nearest Neighbor,” *Build. Informatics, Technol. Sci.*, vol. 3, no. 4, pp. 639–648, 2022, doi: 10.47065/bits.v3i4.1408.
- [12] M. Rizki, D. Devrika, I. H. Umam, and F. S. Lubis, “Aplikasi Data Mining dalam Penentuan Layout Swalayan dengan Menggunakan Metode MBA,” *J. Tek. Ind. J. Has. Penelit. dan Karya Ilm. dalam Bid. Tek. Ind.*, vol. 5, no. 2, p. 130, 2020, doi: 10.24014/jti.v5i2.8958.
- [13] C. Schröer, F. Kruse, and J. M. Gómez, “A systematic literature review on applying CRISP-DM process model,” *Procedia Comput. Sci.*, vol. 181, no. 2019, pp. 526–534, 2021, doi: 10.1016/j.procs.2021.01.199.
- [14] F. N. Dhewayani, D. Amelia, D. N. Alifah, B. N. Sari, and M. Jajuli, “Implementasi K-Means Clustering untuk Pengelompokan Daerah Rawan Bencana Kebakaran Menggunakan Model CRISP-DM,” *J. Teknol. dan Inf.*, vol. 12, no. 1, pp. 64–77, 2022, doi: 10.34010/jati.v12i1.6674.
- [15] A. B. Berek, S. M. Isa, and U. B. Nusantara, “Data Mining untuk Memprediksi Animo Masyarakat terhadap Proses Penerimaan Peserta Didik Baru,” vol. 6, no. 1, pp. 219–226, 2025.
- [16] M. A. Hasanah, S. Soim, and A. S. Handayani, “Implementasi CRISP-DM Model Menggunakan Metode Decision Tree dengan Algoritma CART untuk Prediksi Curah Hujan Berpotensi Banjir,” *J. Appl. Informatics Comput.*, vol. 5, no. 2, pp. 103–108, 2021, doi: 10.30871/jaic.v5i2.3200.
- [17] S. Navisa, Luqman Hakim, and Aulia Nabilah, “Komparasi Algoritma Klasifikasi Genre Musik pada Spotify Menggunakan CRISP-DM,” *J. Sist. Cerdas*, vol. 4, no. 2, pp. 114–125, 2021, doi: 10.37396/jsc.v4i2.162.
- [18] N. Simarmata, Z. A. Nadzir, and D. N. Sari, “Analisis Perubahan Garis Pantai menggunakan Metode Sentinel-1 Dual-Polarized Water Index (SDWI) berbasis Data Multitemporal pada Google Earth Engine,” *Geomatika*, vol. 29, no. 2, pp. 107–120, 2023, [Online]. Available: <https://www.researchgate.net/publication/376397860>
- [19] M. H. P. S. Ikhya Ulummuddin, Anggraini Puspita Sari, “Jurnal Teknologi Terpadu WATERFALL,” vol. 6, no. 22, pp. 72–78, 2020.
- [20] S. Munir, K. B. Seminar, Sudradjat, H. Sukoco, and A. Buono, “The Use of Random Forest Regression for Estimating Leaf Nitrogen Content of Oil Palm Based on Sentinel 1-A Imagery,” *Inf.*, vol. 14, no. 1, 2023, doi: 10.3390/info14010010.
- [21] S. P. UTAMI, “Klasifikasi Kesehatan Mental Usia Remaja Menggunakan Algoritma

- Decision Tree Dan Naive Bayes,” 2024.
- [22] J. Rasiban and S. Praja Raymond Maruli, “Penerapan Data Mining Untuk Memprediksi Penerimaan Peserta Didik Baru,” *Innov. J. Soc. Sci. Res.*, vol. 3, pp. 10065–10079, 2023.
 - [23] F. Rahmasari, M. Rifany, T. Priharyanto, and R. Kurniawan, “Perbandingan Algoritma Machine Learning dalam Klasifikasi Status Banjir di Sumatera Utara,” vol. 2024, no. Senada, pp. 307–318, 2024.
 - [24] H. Celik and A. Cinar, “An Application on Ensemble Learning Using Knime,” *2021 Int. Conf. Data Anal. Bus. Ind. ICDABI 2021*, pp. 400–403, 2021, doi: 10.1109/ICDABI53623.2021.9655815.
 - [25] V. I. Yani, A. Aradea, and H. Mubarok, “Optimasi Prakiraan Cuaca Menggunakan Metode Ensemble pada Naïve Bayes dan C4.5,” *J. Tek. Inform. dan Sist. Inf.*, vol. 8, no. 3, pp. 607–619, 2022, doi: 10.28932/jutisi.v8i3.5455.
 - [26] M. K. Afkar and M. Wali, “Aplikasi Prediksi Produksi Cabai dengan Algoritma C . 45 untuk Dinas Pertanian Provinsi Aceh Berbasis Web Abstrak,” *J. Ilmu Komput. dan Teknol. Inf.*, vol. 1, no. 1, 2024.
 - [27] R. Prasetya, “Penerapan Teknik Data Mining Dengan Algoritma Classification Tree untuk Prediksi Hujan,” *J. Widya Climago*, vol. 2, no. 2, pp. 13–23, 2020.
 - [28] R. K. Rachmansyah and R. Astriratma, “Implementasi Algoritma Extra Trees Untuk Klasifikasi Cuaca Provinsi Dki Jakarta Dengan Oversampling SMOTE,” *Semin. Nas. Mhs. Ilmu Komput. dan Apl.*, pp. 461–472, 2023.

LAMPIRAN

A. Data mentah cuaca harian dan detail stasiun BMKG

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	date	Tn	Tx	Tavg	RH_avg	RR	ss	ff_x	ddd_x	ff_avg	ddd_car	station_id								
2	1/1/2010	21.4	30.2	27.1	82	9	0.5	7	90	5 E	96001									
3	2/1/2010	21	29.6	25.7	95	24	0.2	6	90	4 E	96001									
4	3/1/2010	20.2	26.8	24.5	98	63	0	5	90	4 E	96001									
5	4/1/2010	21	29.2	25.8	90	0	0.1	4	225	3 SW	96001									
6	5/1/2010	21.2	30	26.7	90	2	0.4				96001									
7	6/1/2010	21.2	30	26.1	93	11	0.3				96001									
8	7/1/2010	21.4	29	25.4	96	2	0.1				96001									
9	8/1/2010	21.8	29.8	26.8	91	3	0.6	5	90	4 E	96001									
10	9/1/2010	21.8	30.2	26.7	92	3	0.7	4	90	3 E	96001									
11	10/1/2010	20	30.2	27.1	88	28	0.6	5	90	4 E	96001									
12	11/1/2010	20	30.2	25.7	92	4	0.6	4	90	3 E	96001									
13	12/1/2010	23.4	30.2	27.3	85	0	0.6	3	45	3 NE	96001									
14	13-01-2010	21.8	30.2	26.9	82	0	0.2	4	90	3 E	96001									
15	14-01-2010	22	30	26.3	87	3	0.6	11	90	4 E	96001									
16	15-01-2010	22.4	29.8	26.4	88	1	0.4	8	90	7 E	96001									
17	16-01-2010	22.8	29.8	27.4	82	0	0.6	8	90	7 E	96001									
18	17-01-2010	22.2	30.8	27.5	80	0	0.6	8	90	6 E	96001									
19	18-01-2010	21.8	30.4	26.7	89	0	0.6	4	90	3 E	96001									
20	19-01-2010	20.4	29.8	25.9	95	22	0.2	3	90	3 E	96001									
21	20-01-2010	20.4	29.8	26.1	88	0	0.6	3	90	3 E	96001									
22	21-01-2010	20.2	30	25.5	92	0	0.4				96001									

Lampiran A. 1 *climate_data.csv*

Pada Lampiran A. 1 merupakan tampilan data cuaca harian yang berasal dari platform [kaggle.com](https://www.kaggle.com) yang bersumber dari Badan Meteorologi, Klimatologi, dan Geofisika.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	station_id	station_name	region_name	latitude	longitude	region_id	province_id						
2	96001	Stasiun Meteorologi Maimun Saleh	Kota Sabang	5.87655	95.33785	20	1						
3	96003	Bala Besar Meteorologi Klimatologi dan Geofisika Wilayah III	Kab. Badung	-8.7381	115.1786	272	17						
4	96004	Bala Besar Meteorologi Klimatologi dan Geofisika Wilayah IV	Kota Makassar	-5.14283	119.45227	412	26						
5	96009	Stasiun Meteorologi Malikussaleh	Kab. Aceh Utara	5.22869	96.94749	8	1						
6	96011	Stasiun Meteorologi Sultan Iskandar Muda	Kab. Aceh Besar	5.52244	95.417	6	1						
7	96013	Stasiun Geofisika Aceh Besar	Kab. Aceh Besar	5.496	95.296	6	1						
8	96015	Stasiun Meteorologi Cut Nyak Dhien Nagan Raya	Kab. Nagan Raya	4.04928	96.24796	15	1						
9	96017	Stasiun Klimatologi Aceh	Kab. Aceh Besar	5.404	95.464	6	1						
10	96031	Stasiun Klimatologi Sumatera Utara	Kab. Deli Serdang	3.62114	98.71485	30	2						
11	96033	Stasiun Meteorologi Maritim Belawan	Kota Medan	3.78824	98.71492	49	2						
12	96035	Stasiun Meteorologi Kualanamu	Kab. Deli Serdang	3.64573	98.88488	30	2						
13	96037	Stasiun Geofisika Deli Serdang	Kab. Deli Serdang	3.501	98.56	30	2						
14	96039	Stasiun Geofisika Aceh Selatan	Kab. Aceh Selatan	3.17333	97.28637	1	1						
15	96041	Bala Besar Meteorologi Klimatologi dan Geofisika Wilayah I	Kota Medan	3.5397	98.64	49	2						
16	96043	Stasiun Meteorologi Silangit	Kab. Tapanuli Utara	2.26136	98.99357	25	2						
17	96071	Stasiun Meteorologi Aek Godang	Kab. Padang Lawas Utara	1.55	99.45	43	2						
18	96073	Stasiun Meteorologi FL Tobing	Kab. Tapanuli Tengah	1.55	98.88	24	2						
19	96075	Stasiun Meteorologi Binaka	Kab. Nias	1.1649	97.7036	27	2						
20	96077	Stasiun Geofisika Gunungsitoli	Kab. Nias	1.3	97.58	27	2						
21	96087	Stasiun Meteorologi Hang Nadim	Kota Batam	1.11667	104.11667	150	10						
22	96089	Stasiun Meteorologi Raja Haji Abdullah	Kab. Karimun	1.03	103.38	146	10						

Lampiran A. 2 *station_detail.csv*

Pada Lampiran A. 2 merupakan tampilan data detail stasiun yang berasal dari platform [kaggle.com](https://www.kaggle.com) yang bersumber dari Badan Meteorologi, Klimatologi, dan Geofisika.

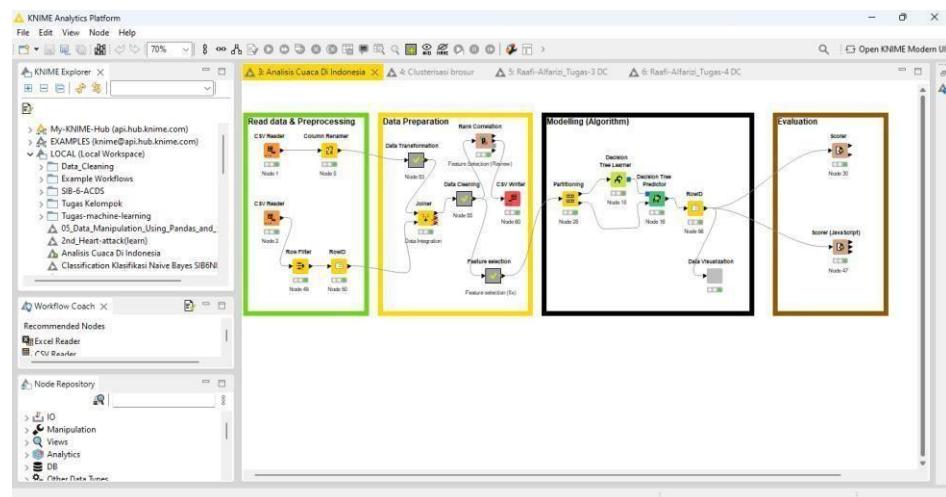
B. Alat Bantu Untuk Analisis Data



© KNIME AG, Zurich, Switzerland - www.knime.com - contact@knime.com

Lampiran B. 1 Aplikasi untuk alat bantu

Pada Lampiran B. 1 yaitu tampilan intro aplikasi yang berguna untuk proses analisis data yang bernama *KNIME Analytics Platform*. Ini adalah perangkat lunak ilmu data Open Source yang memfasilitasi implementasi machine learning dan alat penambangan data, asosiasi data beserta visualisasi data.



Lampiran B. 2 Tampilan isi aplikasi beserta alur kerja

Pada lampiran B. 2 merupakan isi pada aplikasi yang berisi KNIME *Explorer* untuk melihat lembar kerja yang kita buat serta *Node Repository* untuk mengambil *node* sesuai dengan kebutuhan. Selanjutnya terdapat alur kerja/*workflow* beserta *workflow annotation* dari beberapa simpul/*node* yang dipasangkan untuk membaca data, melatih data, pembersihan data, pemodelan data, dan visualisasi data.