LAPORAN RESMI
KEMANAN JARINGAN

Data Mining

Dosen :
Dr. Ferry Astika Saputra ST, M.Sc
Oleh :
Septiana Dyah Anissawati
D4 LJ Teknik Informatika B
3122640031

POLITEKNIK ELEKTRONIKA NEGERI SURABAYA
TAHUN AJARAN
2022/2023

Instalasi Software

Tools yang dibutuhkan :`

1. **Wireshark**

   Wireshark adalah program Network Protocol Analyzer alias penganalisa protokol jaringanyang lengkap. Program ini dapat merakam semua paket yang lewat serta menyeleksi dan menampilkan data tersebut sedetail mungkin.

2. **Knime Analytics Platform**

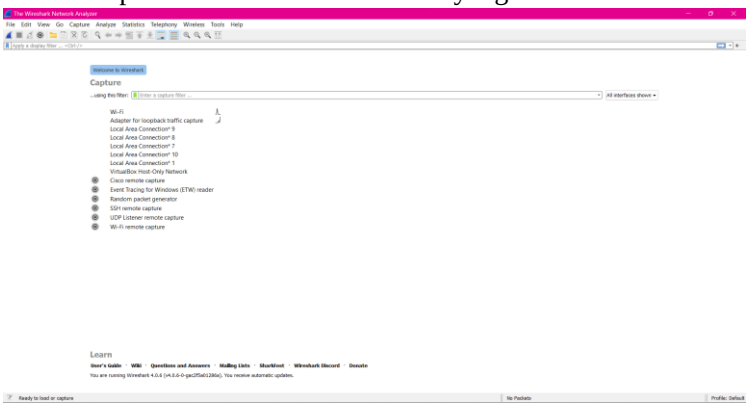
   Knime Analytics Platform adalah software open source untuk membuat model data science. Knime membuat pemahaman data dan merancang alur kerja data science dan komponen yang dapat digunakan kembali.

A. Instalasi Wireshark

a. Software ini dapat didownload pada halaman https://www.wireshark.org/. Kemudian sesuaikan dengan OS pada komputer.

b. Untuk instalasi nya cukup mudah, ikuti saja alur instalasinya dengan pengaturan default.

c. Berikut capture dari software wireshark yang sudah diinstall.

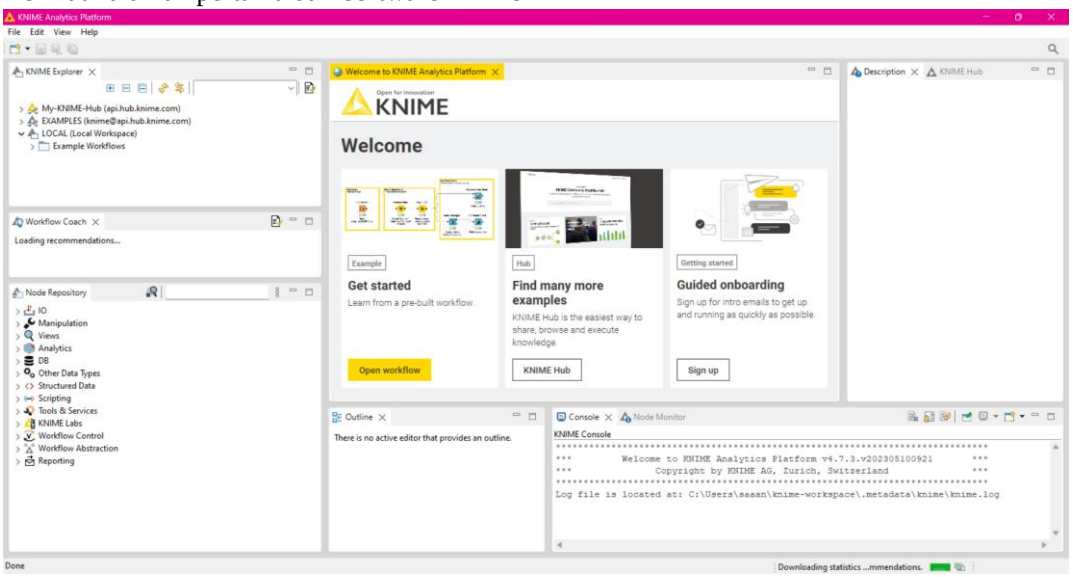


B. Instalasi Knime Analytics Platform

a. Software ini dapat didownload pada halaman https://www.knime.com/downloads/download-knime sesuaikan juga dengan OS pada komputer

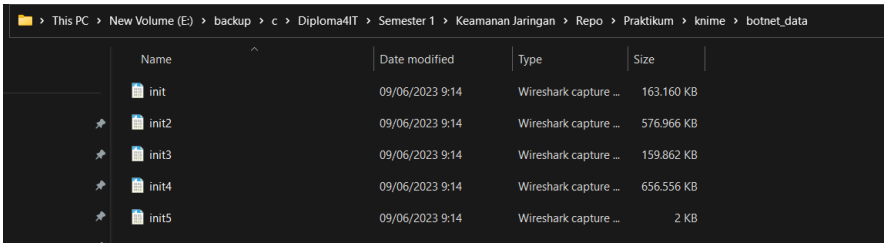
b. Untuk instalasi gunakan pengaturan default pada saat proses instalasi

c. Berikut halaman pertama dari software Knime

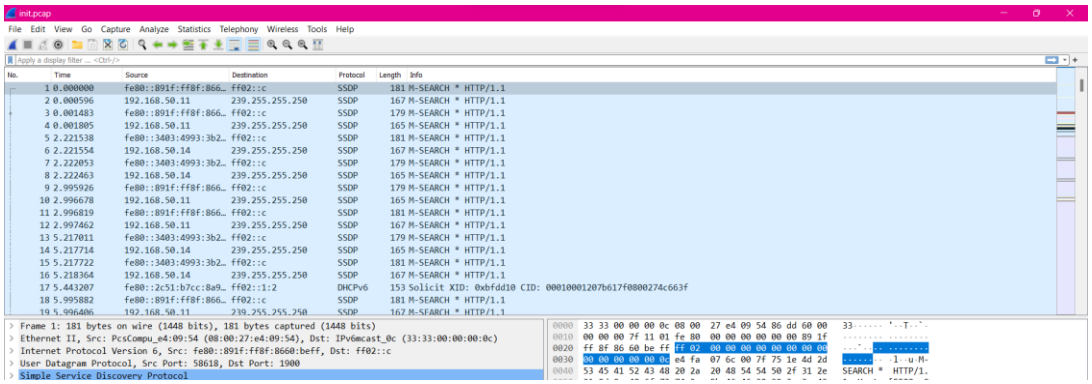


Wireshark

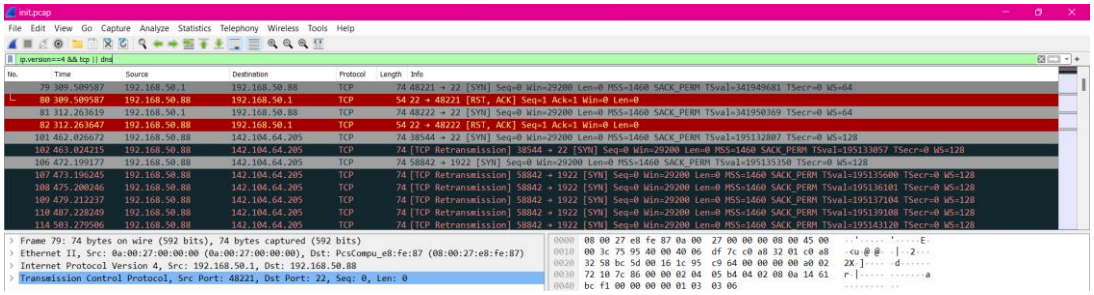
1. Proses ini digunakan untuk mengamati file Packet Capture (.pcap). File tersebut berisi lalu lintas jaringan yang ditangkap oleh komputer. Pada kasus kali ini akan digunakan dataset traffic DNS ISOT yang berasal dari University of Victoria karena terdapat simulasi serangan Botnet pada traffic DNS. Untuk memperoleh dataset ini dapat mengunjungi link :
   https://www.uvic.ca/engineering/ece/isot/datasets/
2. File tersebut dibagi menjadi 5 yaitu : init.pcap, init2.pcap, init3.pcap, init4.pcap, init5.pcap



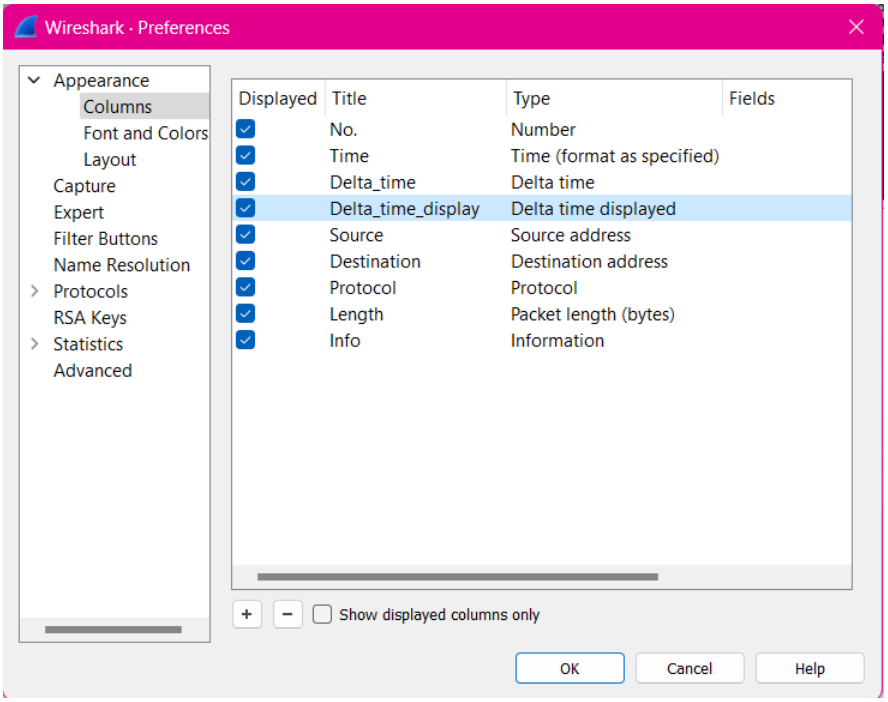
3. Kemudian buka file tersebut secara bergantian menggunakan Wireshark. Pada langkah ini kita gunakan file init.pcap



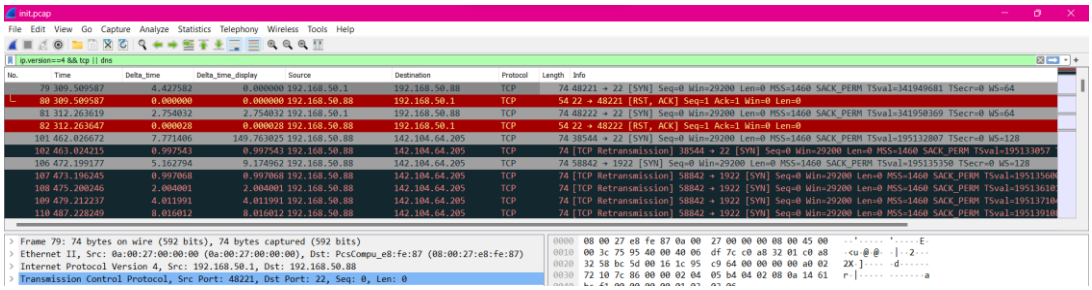
4. Untuk mempermudah pada saat proses analisa yang akan dilakukan nantinya, kita akan mengambil data dengan ip versi 4 (ipv4) dan protocol TCP, DNS saja. Untuk proses tersebut dapat dilakukan pada wireshark menggunakan perintah *ip.version==4 && tcp || dns* pada kolom *display filter* tepat dibawah toolbar.



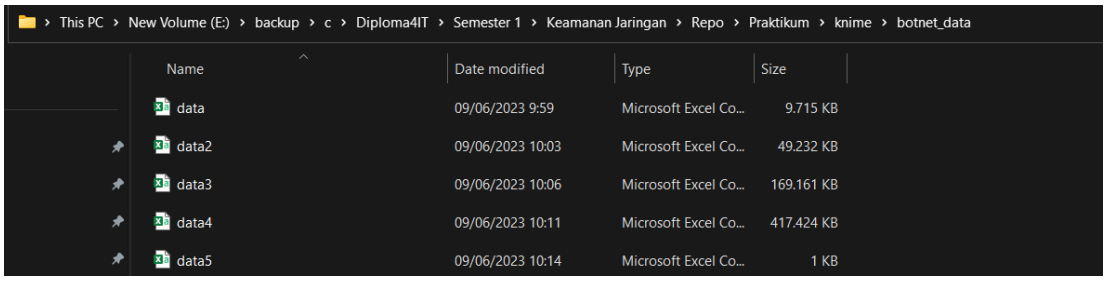
5. Kemudian kita membutuhkan kolom tambahan yaitu delta time. Untuk mendapatkan delta time dan delta time dan delta time display, klik Edit – Preferences – Column



Kemudain klik pada tanda + untuk menambah kolom baru. Kemudian pada Type, pilih Delta Time. Kemudian lakukan hal yang sama untuk kolom delta time display.Kemudian Klik OK. Berikut hasilnya.



6. Langkah terakhir yaitu export file pcap tersebut keformat Comma-separated Value (.csv) dengan cara klik File – Export Packet Dissections – As CSV. Yang perlu diperhatikan yaitu pada Pacet Range, pastikan yang terpilih yaitu Displayed, karena data pada Displayed ini sudah terfilter denga nip version 4.
7. Lakukan semua proses diatas pada dataset berikutnya (init2.pcap, init3.pcap, init4.pcap, init5.pcap)hingga seluruh data sudah terkonversi ke dalam format .csv

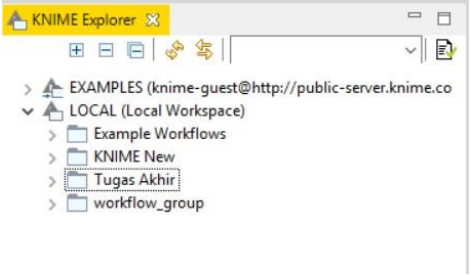


Knime Analytics Platform

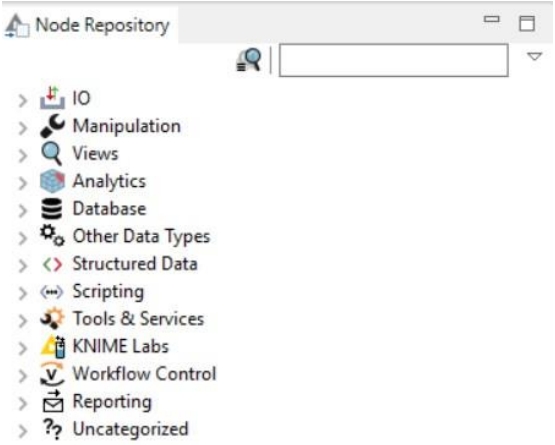
A. Penggabungan Data

1. Setelah semua file tadi telah diexport menjadi file .csv. Buka software Knime Analytics Platform untuk melakukan proses analisa pada traffic DNS. Berikut adalah file yang telah terexport menjadicsv

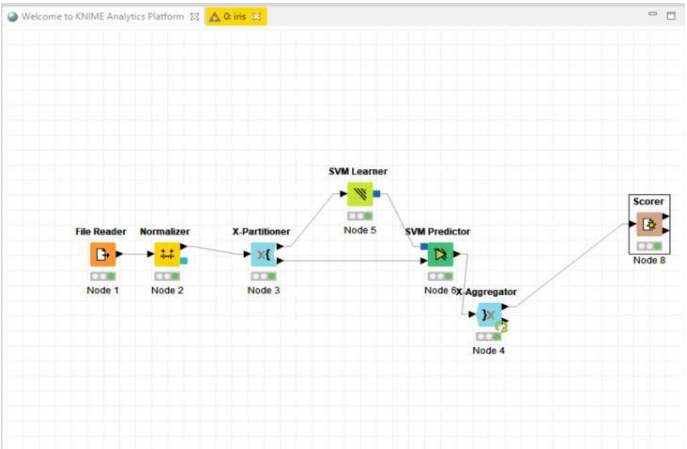
2. Setelah software Knime telah terbuka. Terdapat 3 bagian utama dari software ini. Yang pertama yaitu Knime Explorer yang isinya adalah project project yang kita buat pada software ini.



Kemudian terdapat Node Repository, bagian ini merupakan bagian yang sangat penting, karena berisi seluruh fungsi tools dari software ini yang dinamakan dengan **Node**.



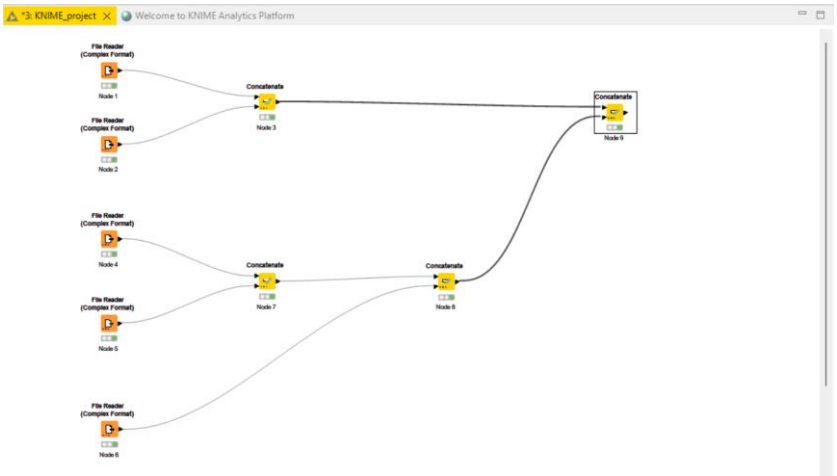
Terakhir yaitu Knime Workflow, bagian ini adalah bagian visual pada Knime, seluruh fungsi yang digunakan akan ditampilkan pada bagian ini. Berikut adalah contoh tampilan pada Knime Workflow



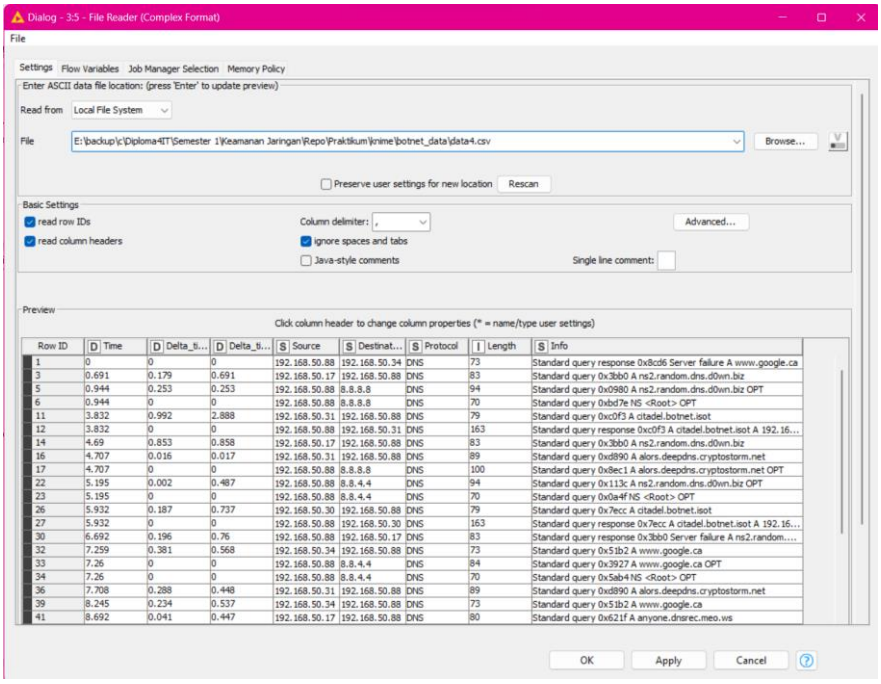
3. Setelah mengenal semua bagian dari Knime. Setelah itu kita akan membuat workflow/project baru. Dengan cara klik File – New – New Knime Workflow – Tulis Nama workflow dan Lokasi workflow tersebut – Klik Finish

4. Selanjutnya yaitu menggabungkan seluruh data tadi menjadi 1 data. Node yang dibutuhkan untukproses ini yaitu :
   a. File Reader : untuk membaca data
   b. Concatenate : untuk menggabungkan data

   Karena node Concatenate hanya dapat menerima input dari 2 data, maka diperlukan lebih dari 1node Concatenate.

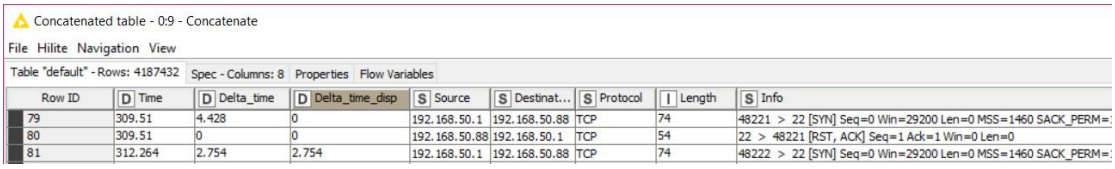


Untuk melihat konfigurasi dari File Reader, dapat digunakan cara klik kanan pada Node, lalu configure



Untuk konfigurasi file reader hanya tinggal memasukkan file csv yang telah diexport pada langkah sebelumya. Klik Aplly – OK. Proses ini belum selesai, karena Node belum di jalankan, untuk

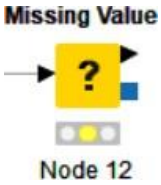
menjalankan Node bisa dengan cara klik kanan pada Node – Execute. Bila berhasil dijalankan, status Node yang berada dibawah Node akan berubah berwarna Hijau.



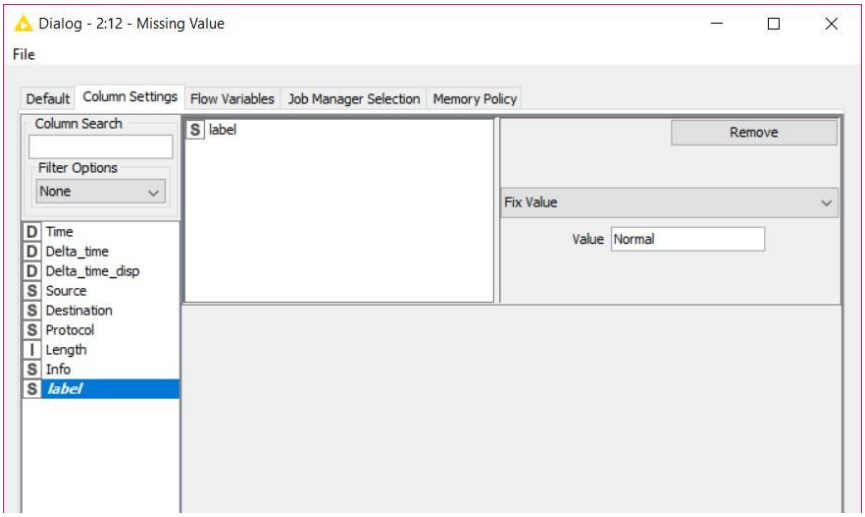
Row ID	Time	Delta_time	Delta_time_disp	Source	Destinat...	Protocol	Length	Info
79	309.51	4.428	0	192.168.50.1	192.168.50.88	TCP	74	48221 > 22 [SYN] Seq=0 Win=29200 Len=0 MSS=1460 SACK_PERM=
80	309.51	0	0	192.168.50.88	192.168.50.1	TCP	54	22 > 48221 [RST, ACK] Seq=1 Ack=1 Win=0 Len=0
81	312.264	2.754	2.754	192.168.50.1	192.168.50.88	TCP	74	48222 > 22 [SYN] Seq=0 Win=29200 Len=0 MSS=1460 SACK_PERM=

Pelabelan Data

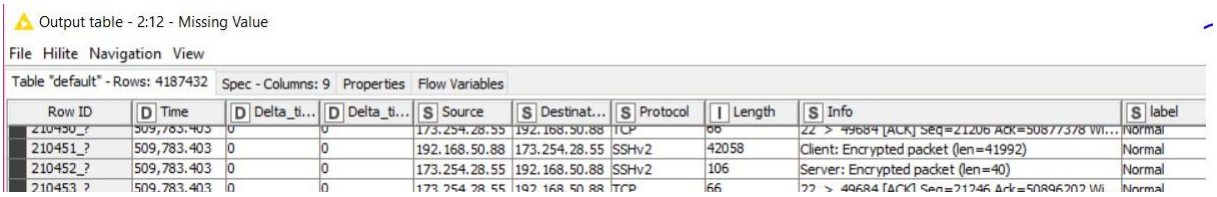
1. Dari langkah 7 menghasilkan output data dengan label malicious tetapi masih terdapat data dengan berlabel "**?**". Hal ini dikarenakan kita hanya labeling untuk data malicious saja. Untuk melakukan labeling data normal kita akan menggunakan Node **Missing Value.** Node ini digunakanuntuk mengisi data kosong.



Untuk konfigurasi, dapat menggunakan konfigurasi berikut.

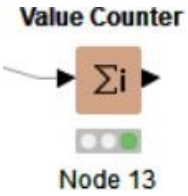


Konfigurasi ini nantinya akan mengisi value yang kosong dengan value Normal. Berikut hasil dariproses Missing Value



Row ID	Time	Delta_ti...	Delta_ti...	Source	Destinat...	Protocol	Length	Info	label
210450_?	509,783.403	0	0	173.254.28.55	192.168.50.88	TCP	66	22 > 49684 [ACK] Seq=21206 Ack=50877378 Wi...	Normal
210451_?	509,783.403	0	0	192.168.50.88	173.254.28.55	SSHv2	42058	Client: Encrypted packet (len=41992)	Normal
210452_?	509,783.403	0	0	173.254.28.55	192.168.50.88	SSHv2	106	Server: Encrypted packet (len=40)	Normal
210453_?	509,783.403	0	0	173.254.28.55	192.168.50.88	TCP	66	22 > 49684 [ACK] Seq=21246 Ack=50896202 Wi...	Normal

- Untuk memastikan bahwa kolom label sudah terisi dengan value Malicious atau Normal, dapat menggunakan node **Value Counter.** Node ini berfungsi untuk menghitung jumlah seluruh value pada kolom terpilih.

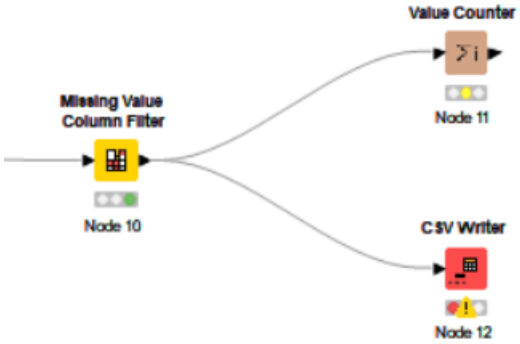


Berikut adalah hasil dari perhitungan value dengan Value Counter. Dalam konfigurasinya tinggalmemilih kolom yang ingin dihitung yaitu kolom label.

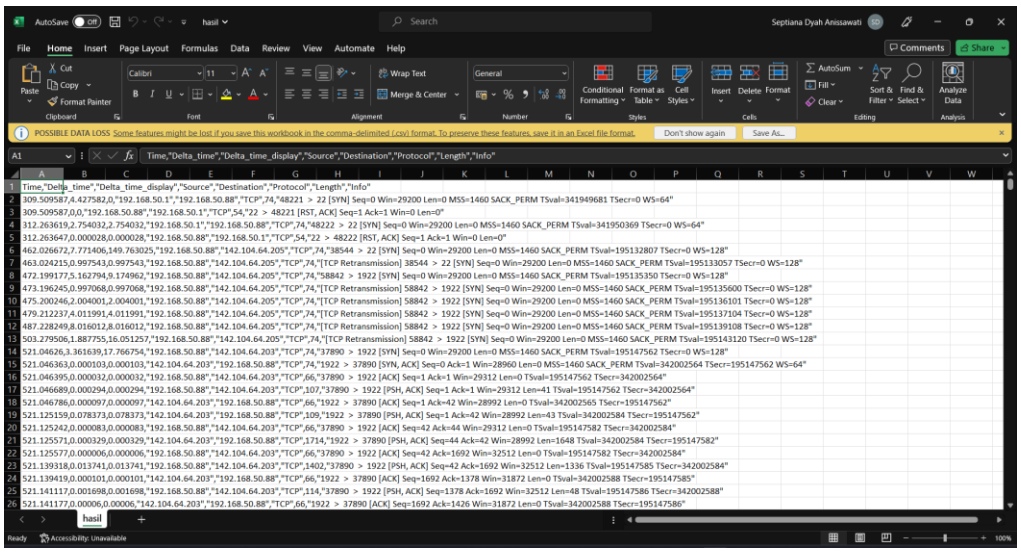
Row ID	count
Normal	3135405
malicious	1052027

Dari gambar tersebut bisa dilihat jika sudah tidak ada data yang memiliki label kosong, hanya terdapat 2 label yaitu Normal dan malicious

- Export file ke dalam format .csv dengan menggunakan node **CSV Writer**



- Export file dan hasil



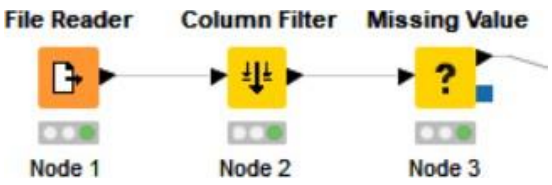
Knime For Data Mining

Pada langkah ini, kita akan melakukan sebuah analisa pada data ISOT dengan menggunakan platform Knime. Pada proses sebelumnya kita sudah melakukan penggabungan data dan pelabelan data. Pada proses ini kita akan melakukan analisa data dengan menggunakan Teknik **Data Mining.** Data mining adalah proses penggalian data atau penemuan informasi baru dengan mencari pola atau aturan tertentu dari sebuah data. Di dalam Data Mining terdapat beberapa proses yang harus dilakukan terlebih dahulu sebelum dilakukannya data mining. Berikut tahapan-tahapan untuk melakukan data mining, yaitu

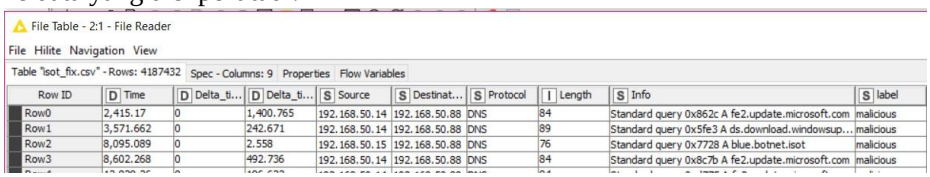
1. Data Pre-Processing
2. Data Transformation
3. Data Mining
4. Evaluation

Data Pre-processing

1. Proses ini adalah proses dimana data akan dibersihkan (cleaning) karena biasanya didalam suatu data terdapat nilai-nilai yang tidak sempurna atau bahkan terdapat nilai-nilai yang hilang atau kosong yang nantinya akan dapat mempengaruhi proses kedepannya. Pada proses ini kita membutuhkan Node-node berikut : File Reader, Column Filter, Missing Value.

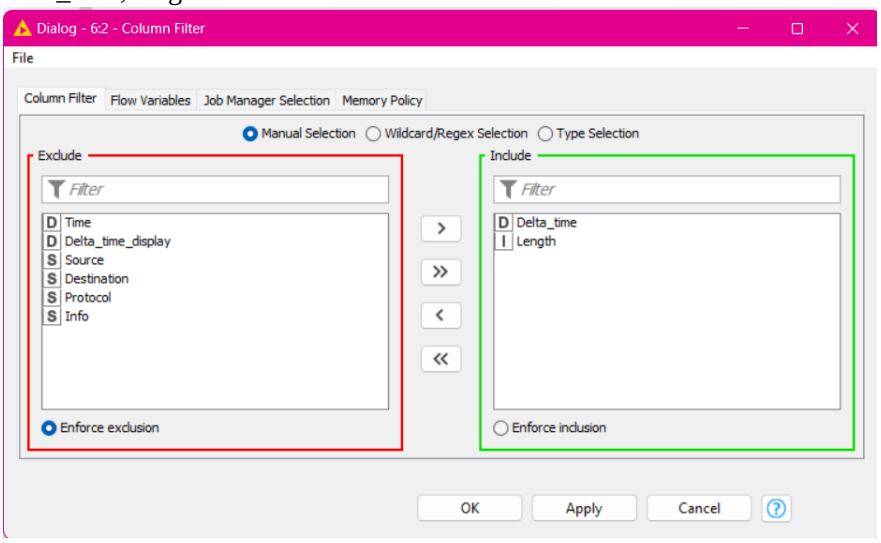


2. Sebelum memulai data pre-processing, langkah pertama yaitu membaca file data csv yang sudah diexport pada proses sebelumnya menggunakan Node File Reader. Untuk konfigurasinya cari lokasi file data yang diexport tadi.

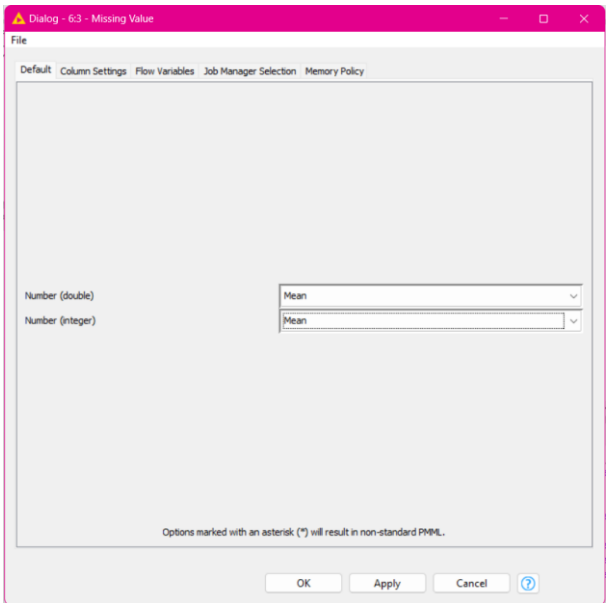


Row ID	Time	Delta_ti...	Delta_ti...	Source	Destinat...	Protocol	Length	Info	label
Row0	2,415.17	0	1,400.765	192.168.50.14	192.168.50.88	DNS	84	Standard query 0x862c A fe2.update.microsoft.com	malicious
Row1	3,571.662	0	242.671	192.168.50.14	192.168.50.88	DNS	89	Standard query 0x5fe3 A ds.download.windowsup...	malicious
Row2	8,095.089	0	2.558	192.168.50.15	192.168.50.88	DNS	76	Standard query 0x7728 A blue.botnet.isot	malicious
Row3	8,602.268	0	492.736	192.168.50.14	192.168.50.88	DNS	84	Standard query 0x8c7b A fe2.update.microsoft.com	malicious

3. Selanjutnya kita akan menggunakan Node **Column Filter.** Node ini berfungsi untuk mem-filter kolom atau atribut yang tidak digunakan. Disini kita hanya menggunakan 3 atribut, yaitu : Delta_time, length dan label.

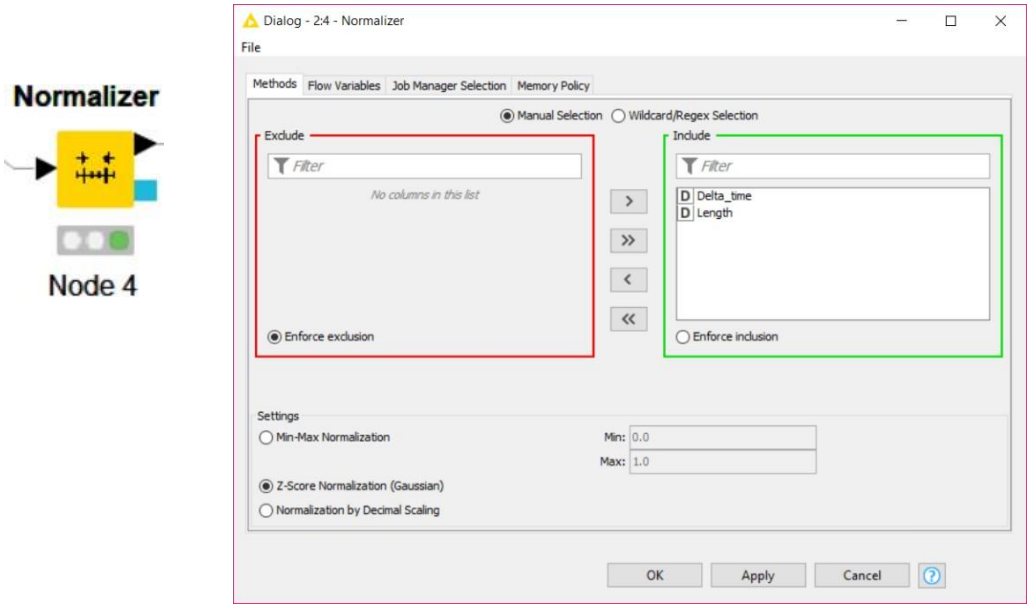


4. Setelah langkah 2 dijalankan, data tersebut menjadi memiliki 3 kolom atau atribut yang sebelumnya terdapat 9 kolom. Selanjutnya kita akan menjalankan Node **Missing Value**. Node ini sudah pernah kita pakai pada proses labeling data. Tetapi pada proses ini kita akan melakukan pembersihan data, karena biasanya didalam suatu data terdapat kolom yang tidak sempurna seperti data yang hilang atau atribut yang tidak relevan, untuk itu Node ini diperlukan untuk mengatasi hal tersebut. Berikut konfigurasinya

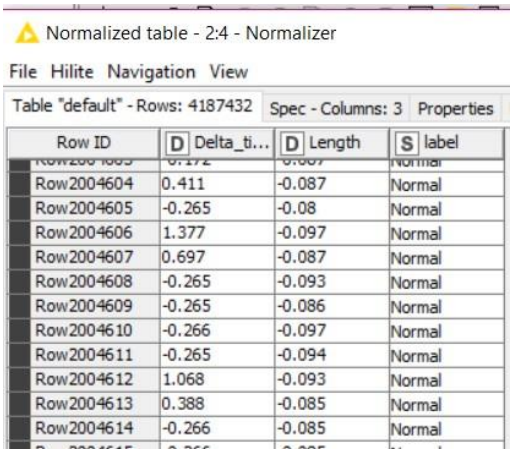


Data Transformation

1. Setelah melakukan data pre-processing, selanjutnya akan menuju ke proses data transformation, pada proses ini data akan diubah ke format yang sesuai untuk proses data mining. Node yang diguanakn pada tahap ini yaitu **Normalizer**. Berikut konfigurasinya



a. Setelah dijalankan, kolom dari data tersebut akan berubah menjadi bentuk range. Data inilah yang nantinya akan digunakan dalam pengenalan pola. Berikut adalah hasil dari Node Normalizer



Normalized table - 2:4 - Normalizer

File Hilite Navigation View

Table "default" - Rows: 4187432 Spec - Columns: 3 Properties

Row ID	D Delta_ti...	D Length	S label
Row2004604	0.411	-0.087	Normal
Row2004605	-0.265	-0.08	Normal
Row2004606	1.377	-0.097	Normal
Row2004607	0.697	-0.087	Normal
Row2004608	-0.265	-0.093	Normal
Row2004609	-0.265	-0.086	Normal
Row2004610	-0.266	-0.097	Normal
Row2004611	-0.265	-0.094	Normal
Row2004612	1.068	-0.093	Normal
Row2004613	0.388	-0.085	Normal
Row2004614	-0.266	-0.085	Normal

Setelah mendapatkan data ini, baru kita dapat menjalankan proses Data Mining

Data Mining

b. Setelah menyelesaikan tahap data transformation, kita akan menjalankan proses Data Mining, dalam proses ini kita akan menggunakan Metode Klasifikasi **Decision Tree** dengan teknik **Cross Validation**. Pada proses ini kita membutuhkan Node-node berikut : X-Partitioner, Decision Tree Learner, Decision Tree Predictor, X-Aggregator. Sehingga akan membentuk flow seperti ini

c. X-Partitioner berfungsi untuk menentukan jumlah iterasi atau pengulangan pada teknik cross validation, data ini nantinya akan terbagi menjadi 2 yaitu data training dan data testing. Berikut konfigurasinya

d. Decision Tree Learner berfungsi sebagai data training, karena metode Decision Tree merupakan supervised learning, sehingga membutuhkan data training untuk mengenali pola dari setiap data.Berikut konfigurasi dari Decision Tree Learner

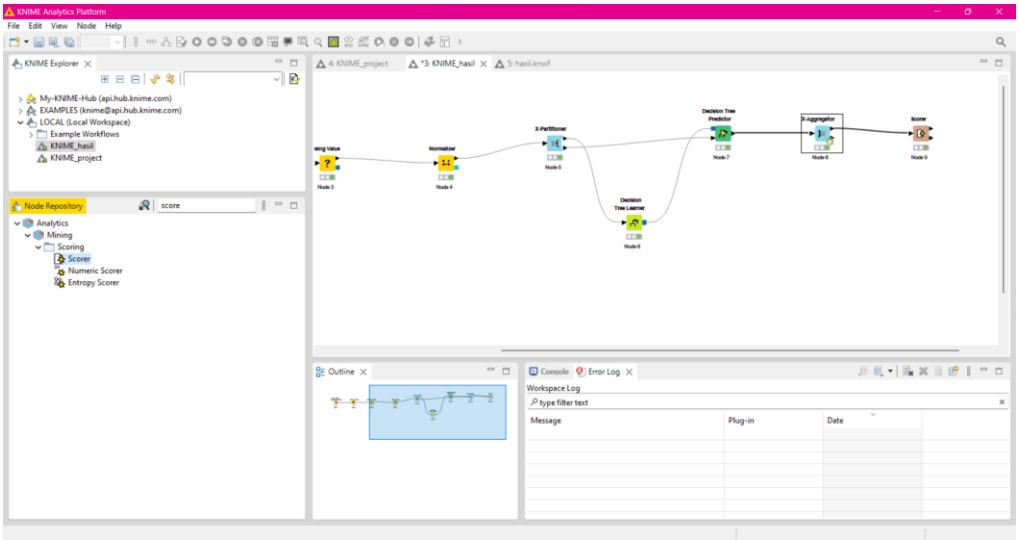
e. Setelah menjalankan Decision Tree Learner, lalu dilanjutkan dengan Decision Tree Predictor. Node ini berfungsi untuk menklasifikasi data dengan cara menguji data testing dengan hasil dari prosesDecision Tree Learner. Berikut konfigurasiny

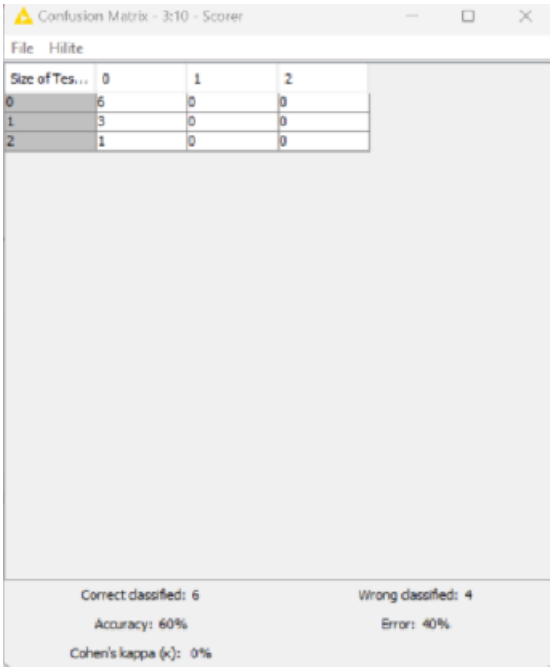
f. Node X-Aggregator berfungsi sebagai akhir dari proses cross validation. Node ini akan mengumpulkan hasil dari Node Predictor yang akan menampilkan hasil dari prediksi dari beberapa iterasi yang dilakukan. Tidak ada konfigurasi khusus dari node ini, sehingga bisa langsung dijalankan. Berikut adalah hasil dari node X-Aggregator. Dari hasil ini akan mendapatkan kolom baru yaitu kolom prediksi.

i. Evaluation

1. Proses ini merupakan proses terakhir pada tahap data mining yaitu merupakan hasil dari teknik data mining berupa hasil prediksi untuk menilai apakah model ini dapat digunakan untuk mengenali pola serangan pada data ISOT. Untuk langkah ini kita akan menggunakan Node **Scorer** yang didalamnya terdapat perhitungan untuk melihat seberapa baik model ini dengan menggunakan teknik confusion matrix. Berikut konfigurasinya.

2. Setelah node ini dijalankan, kita dapat melihat presentase dari hasil confusion matrix. Hasil inilah yang nantinya akan digunakan untuk menentukan keputusan apakah model ini baik atau tidak dalam menanagani kasus data ISOT Botnet untuk mendeteksi serangan.




Confusion Matrix - 3:10 - Scorer
File Hilite
Size of Tes... 0 1 2
0 6 0 0
1 3 0 0
2 1 0 0
Correct classified: 6
Wrong classified: 4
Accuracy: 60%
Error: 40%
Cohen's kappa (κ): 0%