

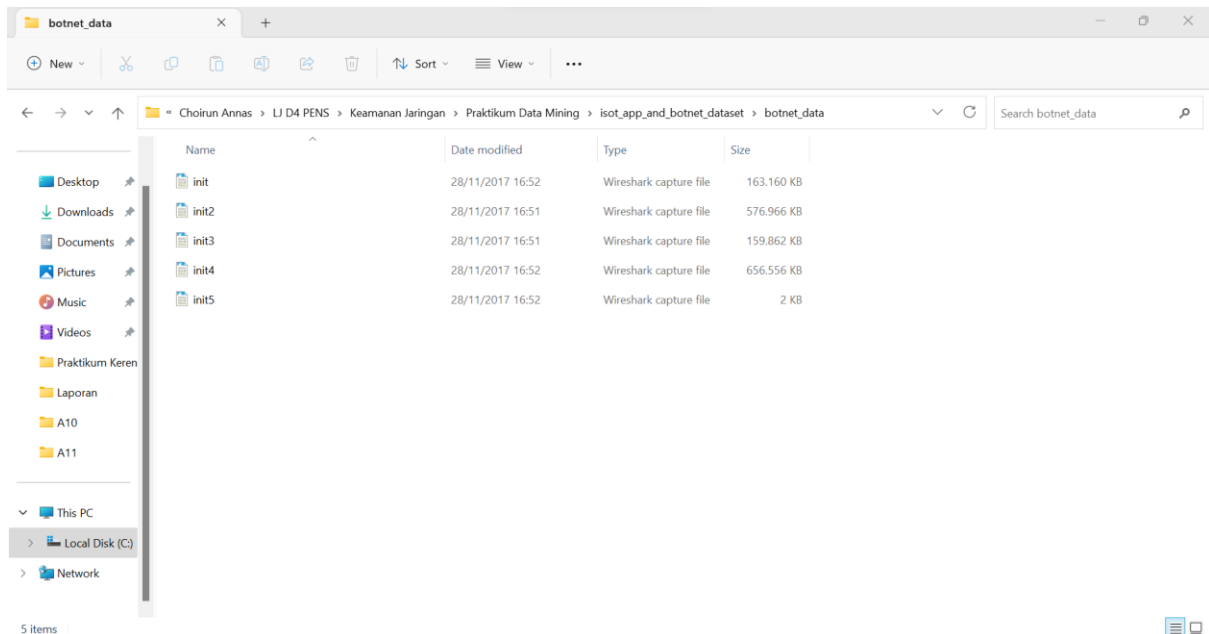
# PRAKTIKUM DATA MINING



Nama	: Choirun Annas
NRP	: 3122640032
Mata Kuliah	: Keamanan Jaringan
Dosen	: Bapak Dr. Ferry Astika Saputra ST, M.Sc

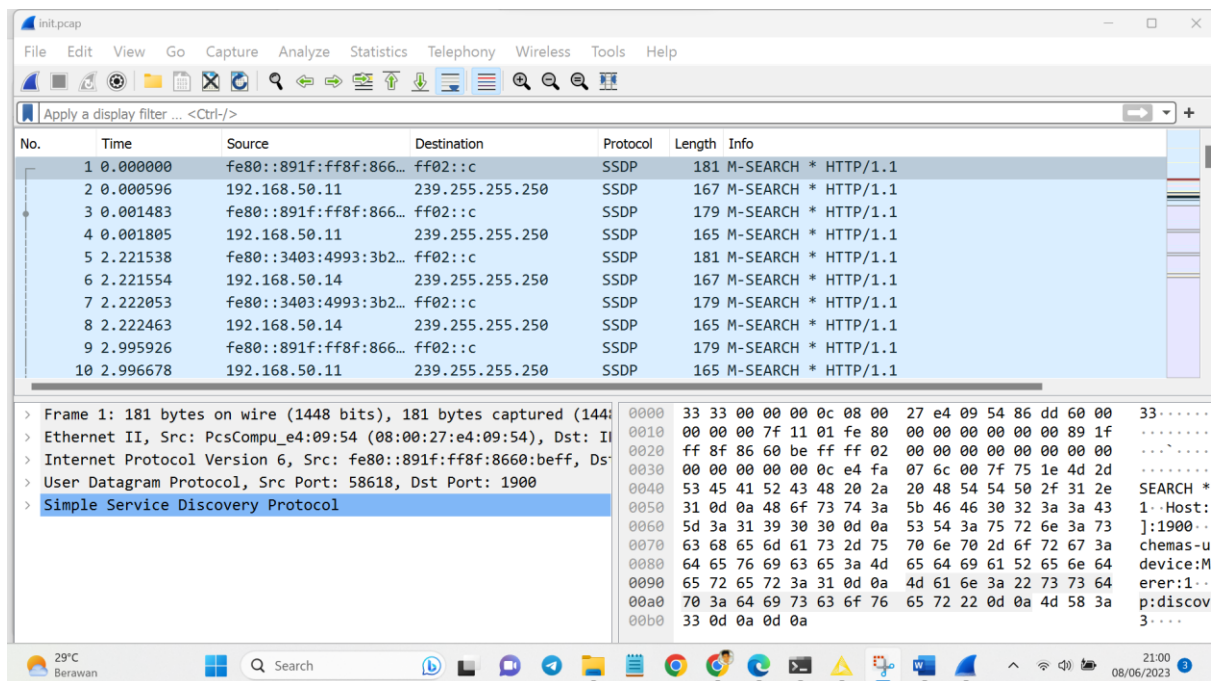
# Laporan Praktikum

1. Bagi File menjadi 5 bagian : init.pcap, init2.pcap, init3.pcap, init4.pcap, init5.pcap

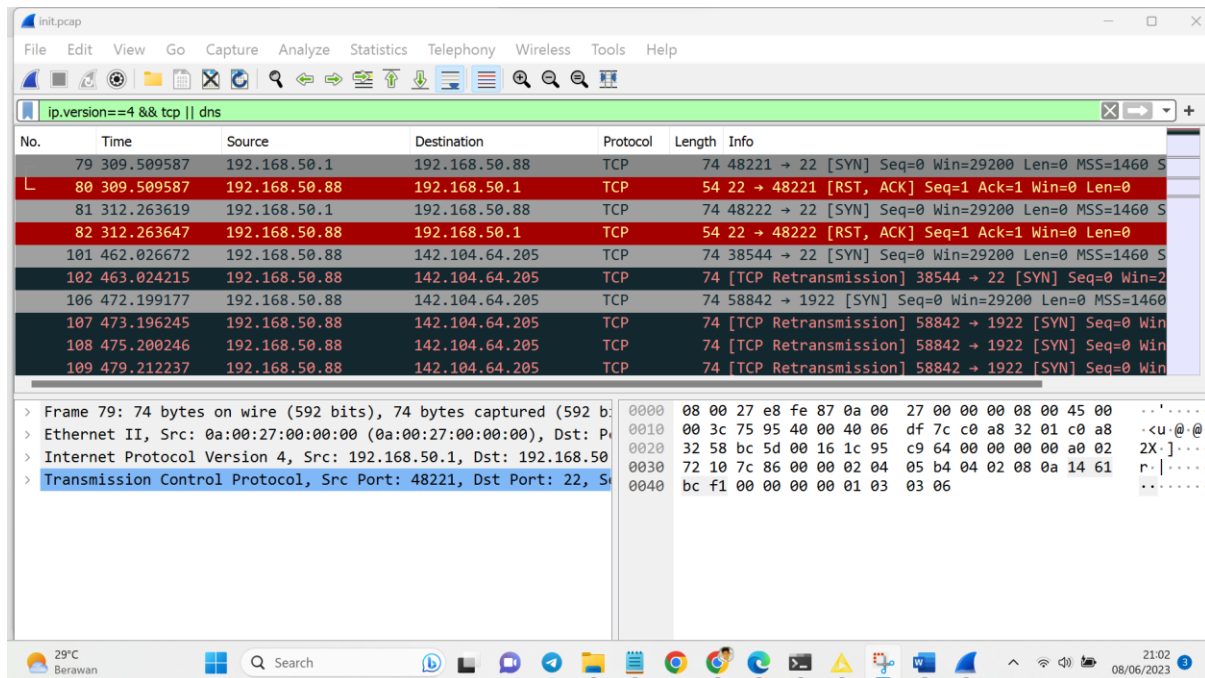


2. Kemudian buka file tersebut secara bergantian menggunakan Wireshark. Pada langkah ini kita gunakan

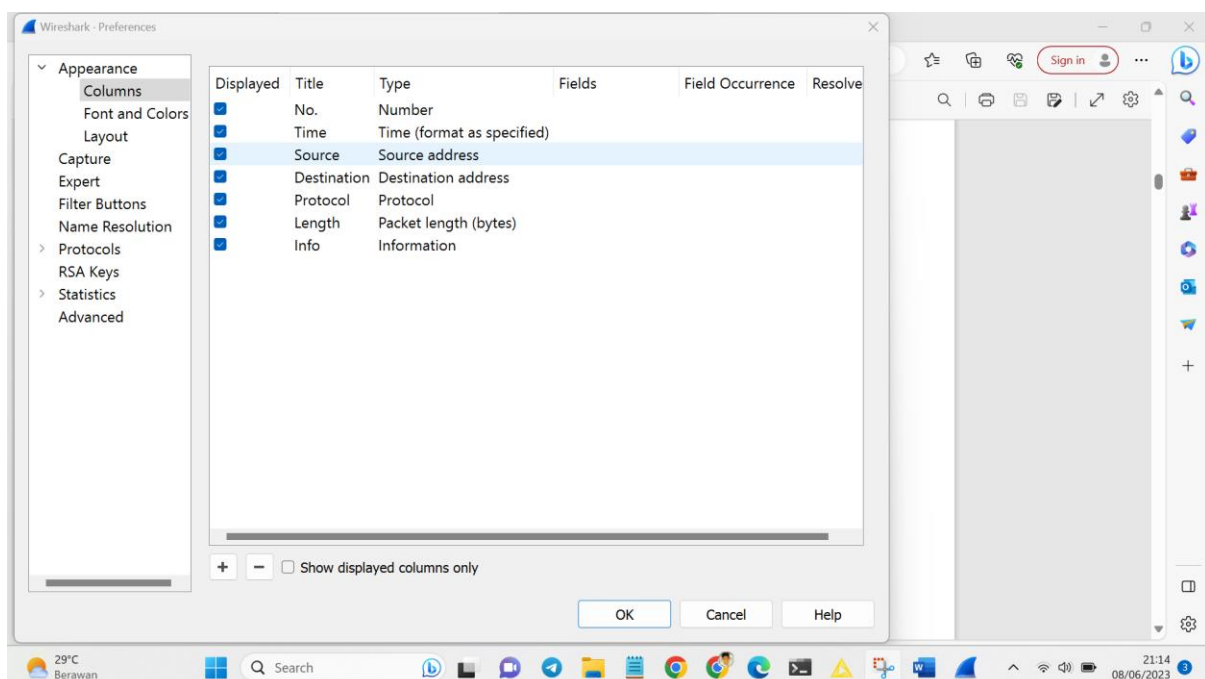
file init.pcap



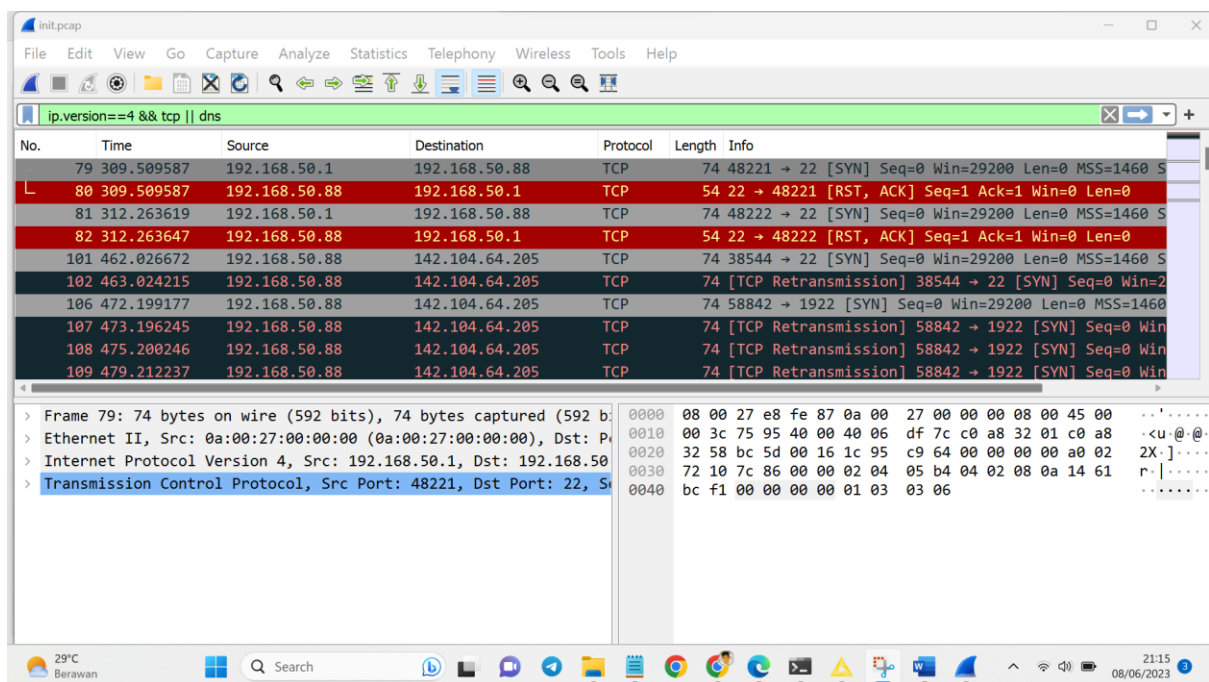
3. Untuk proses analisa yang akan dilakukan nantinya, kita akan mengambil data dengan ip versi 4 (ipv4) dan protocol TCP, DNS saja. Untuk proses tersebut dapat dilakukan pada wireshark menggunakan perintah `ip.version==4 && tcp || dns` pada kolom display filter tepat dibawah toolbar



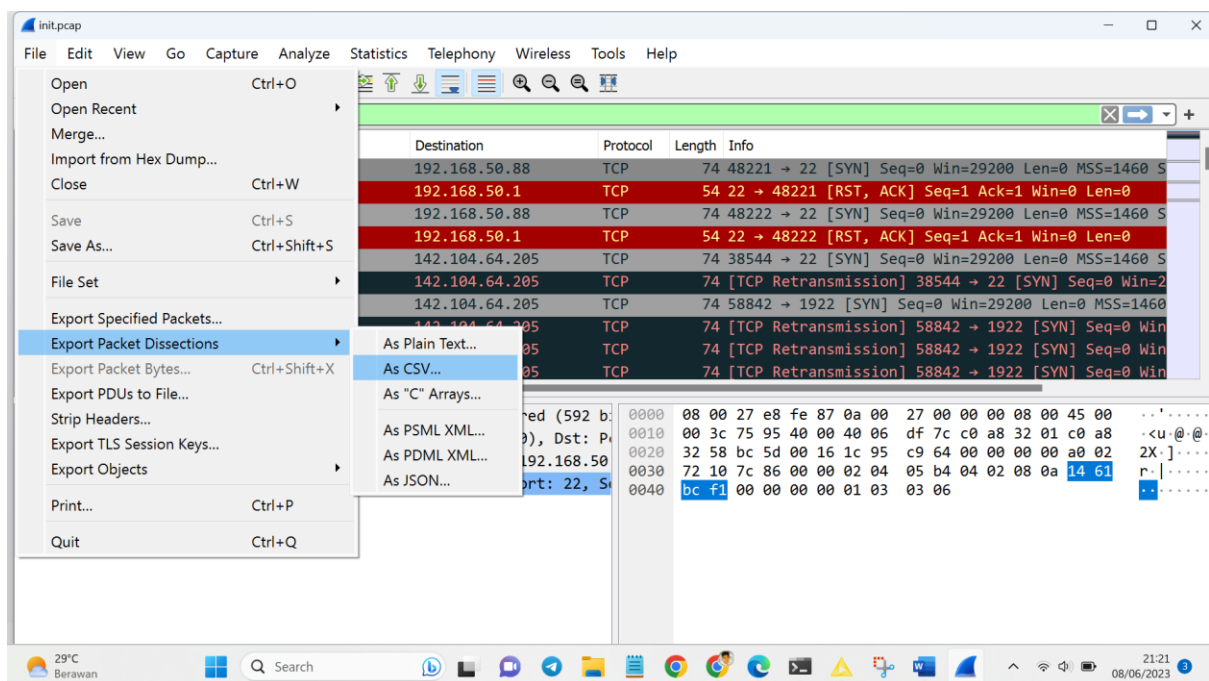
4. Untuk mendapatkan delta time dan delta time display, klik Edit – Preferences – Column

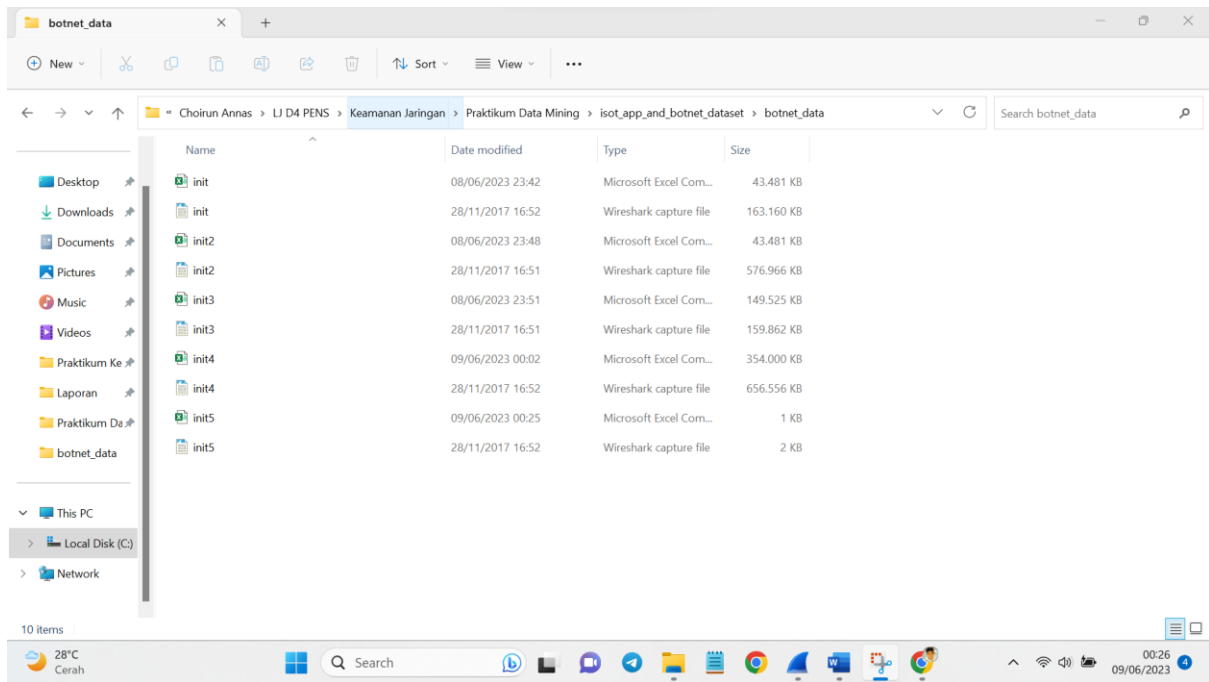


Hasilnya

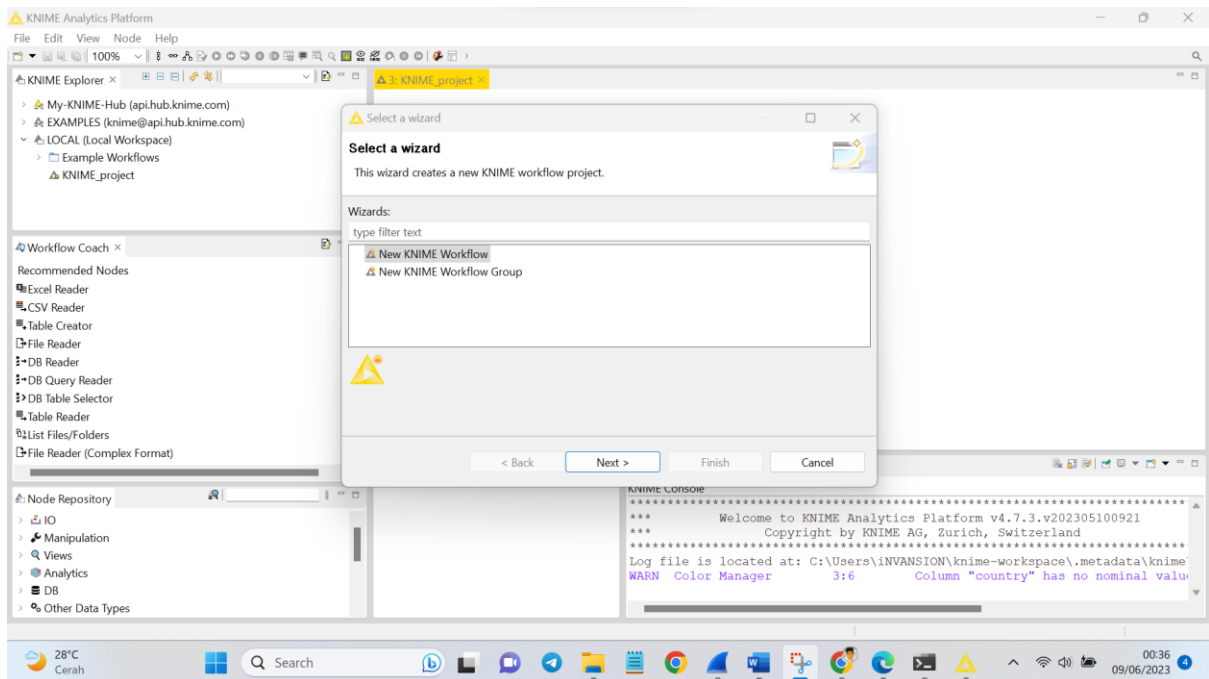


5. Export file pcap tersebut keformat Comma-separated Value (.csv) dengan cara klik File – Export Packet Dissections – As CSV. Yang perlu diperhatikan yaitu pada Pacet Range, pastikan yang terpilih yaitu Displayed, karena data pada Displayed ini sudah terfilter dengan nip version 4

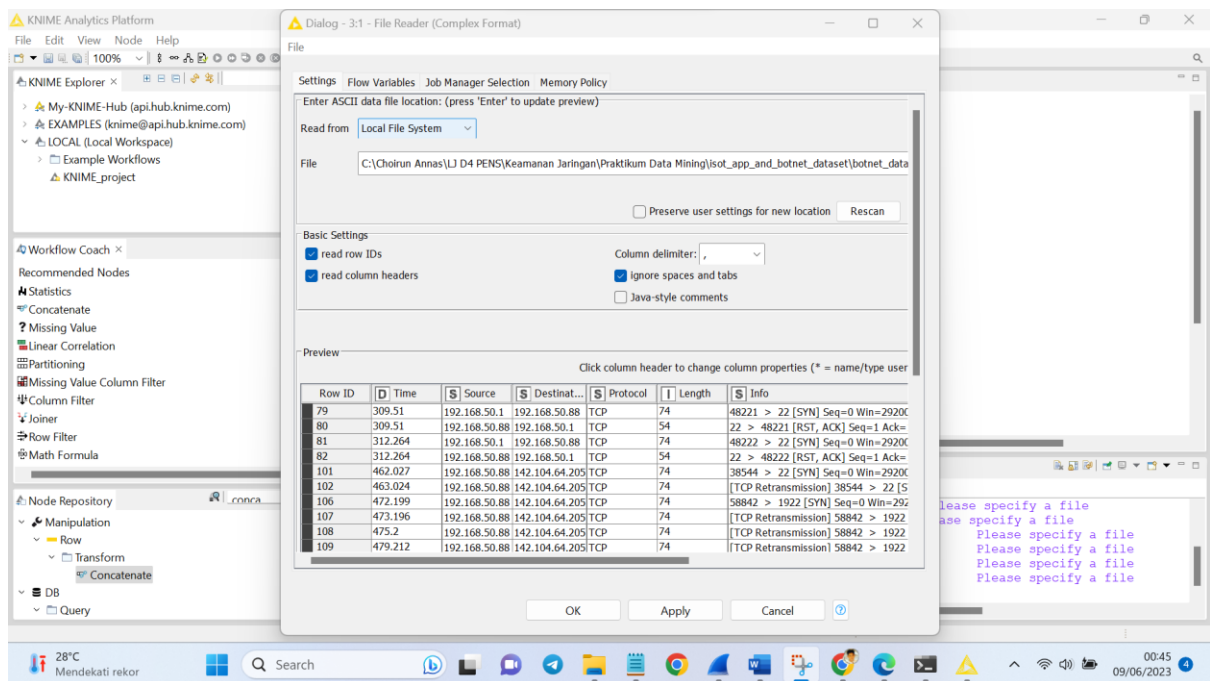




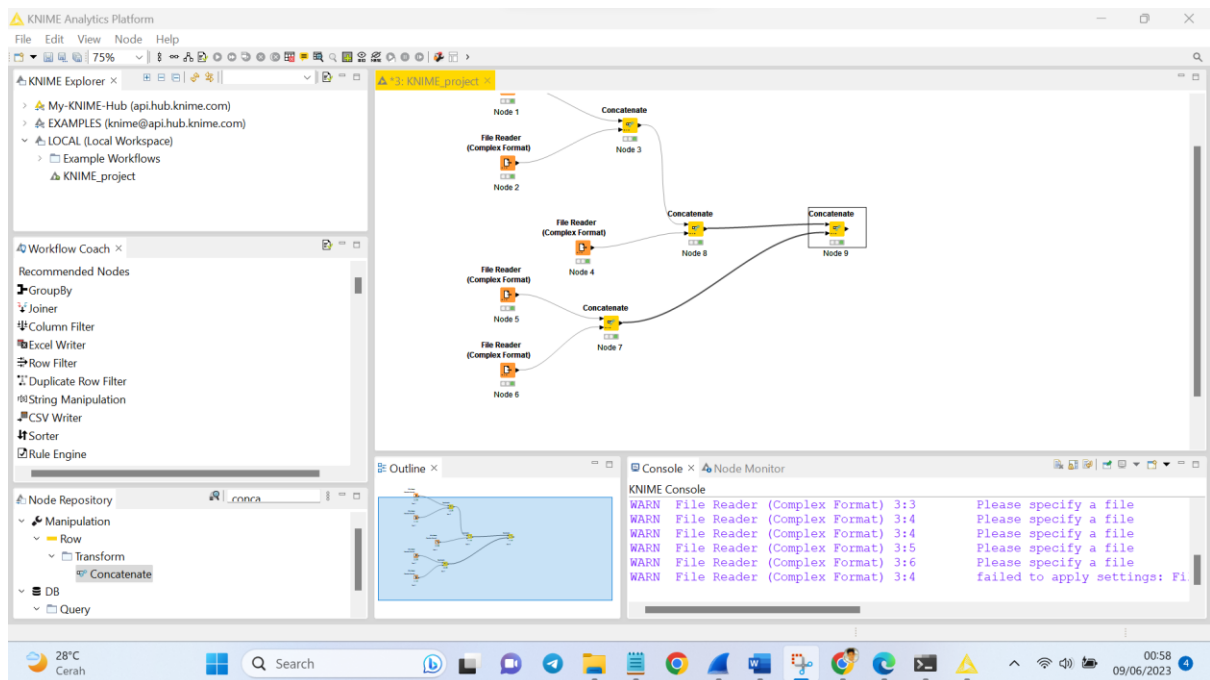
5. membuat workflow/project baru. Dengan cara klik File – New – New Knime Workflow – Tulis Nama workflow dan Lokasi workflow tersebut – Klik Finish



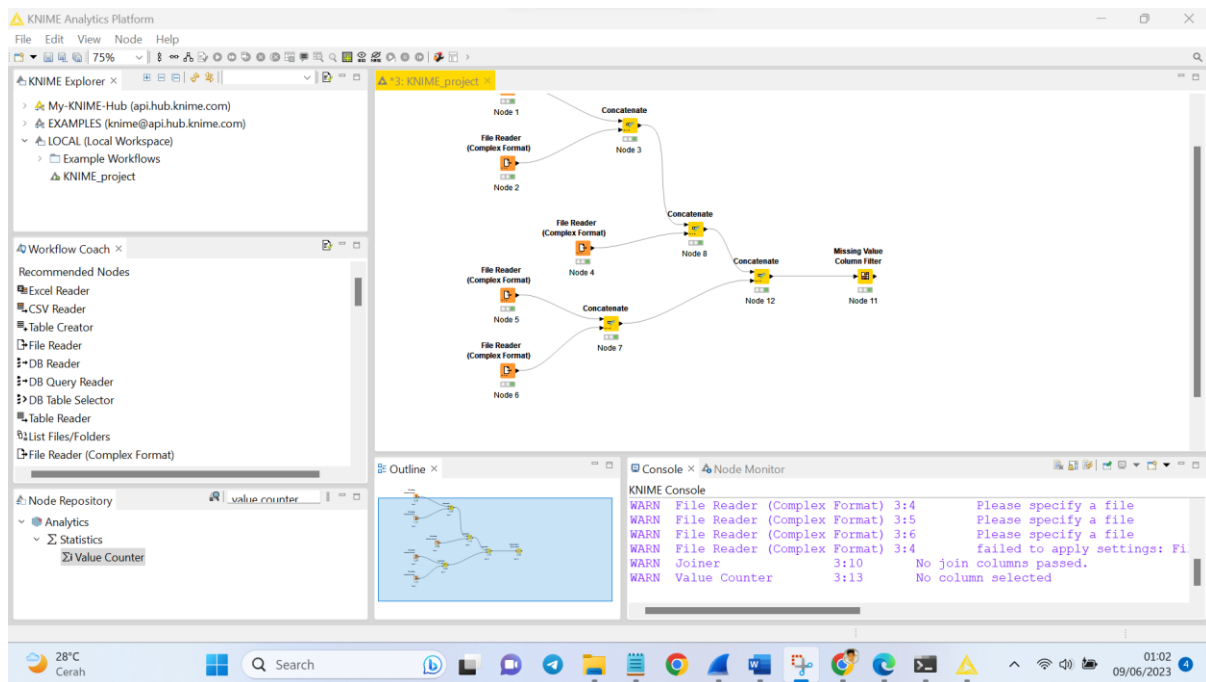
## 6. Tambahkan data kedalam file reader



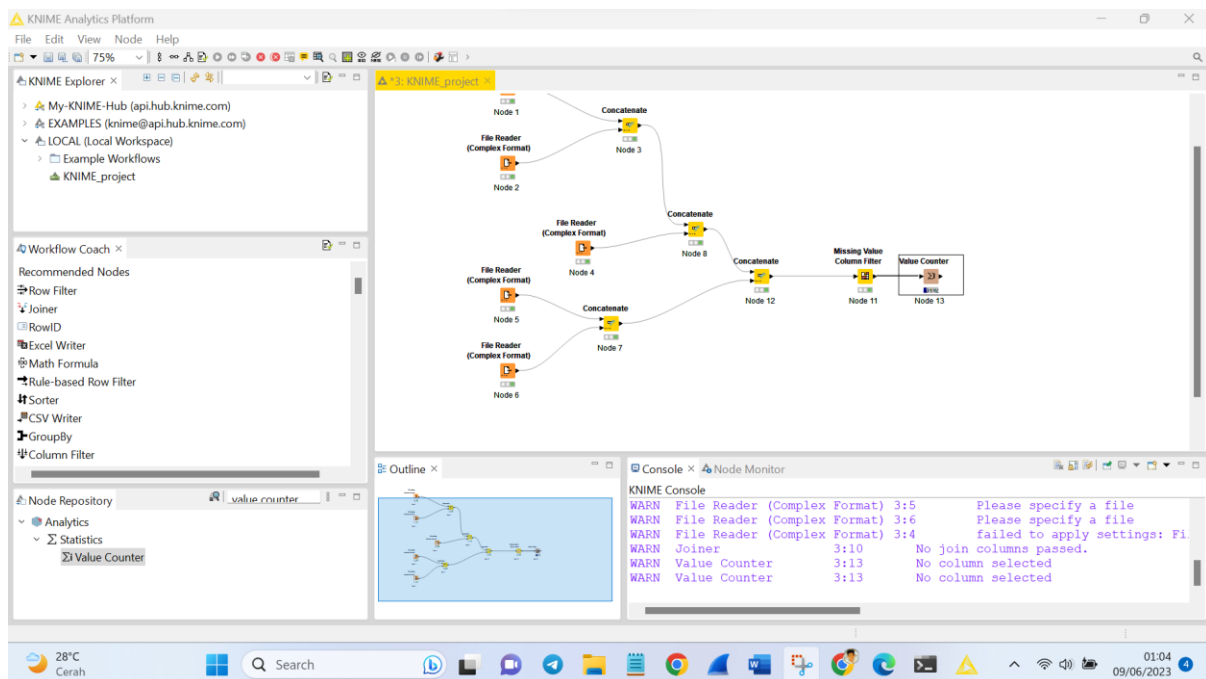
## 7. Gabungkan kelima data dengan menggunakan concatenate dan data reader seperti gambar dibawah



8. Untuk melakukan labeling data normal kita akan menggunakan Node Missing Value. Node ini digunakan untuk mengisi data kosong.

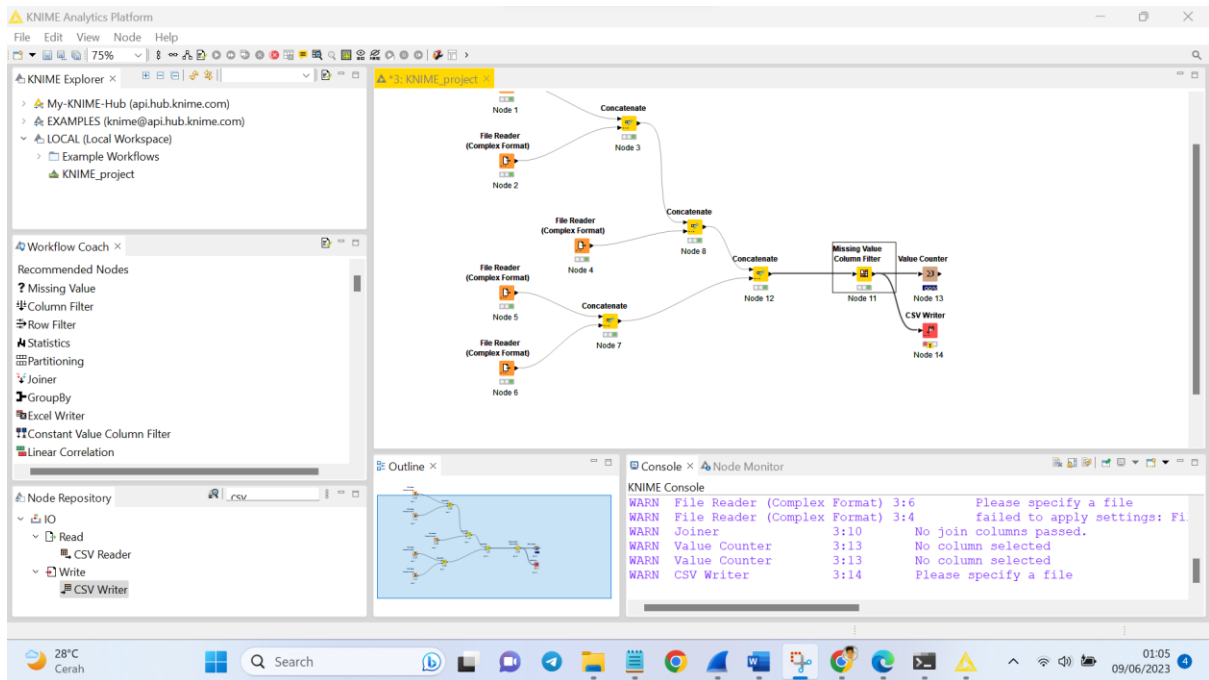


9. Untuk memastikan bahwa kolom label sudah terisi dengan value Malicious atau Normal, dapat menggunakan node Value Counter. Node ini berfungsi untuk menghitung jumlah seluruh value pada kolom terpilih.





## 10. Export file ke dalam format .csv dengan menggunakan node CSV Writer



Saya menaruh data didalam file hasil.csv

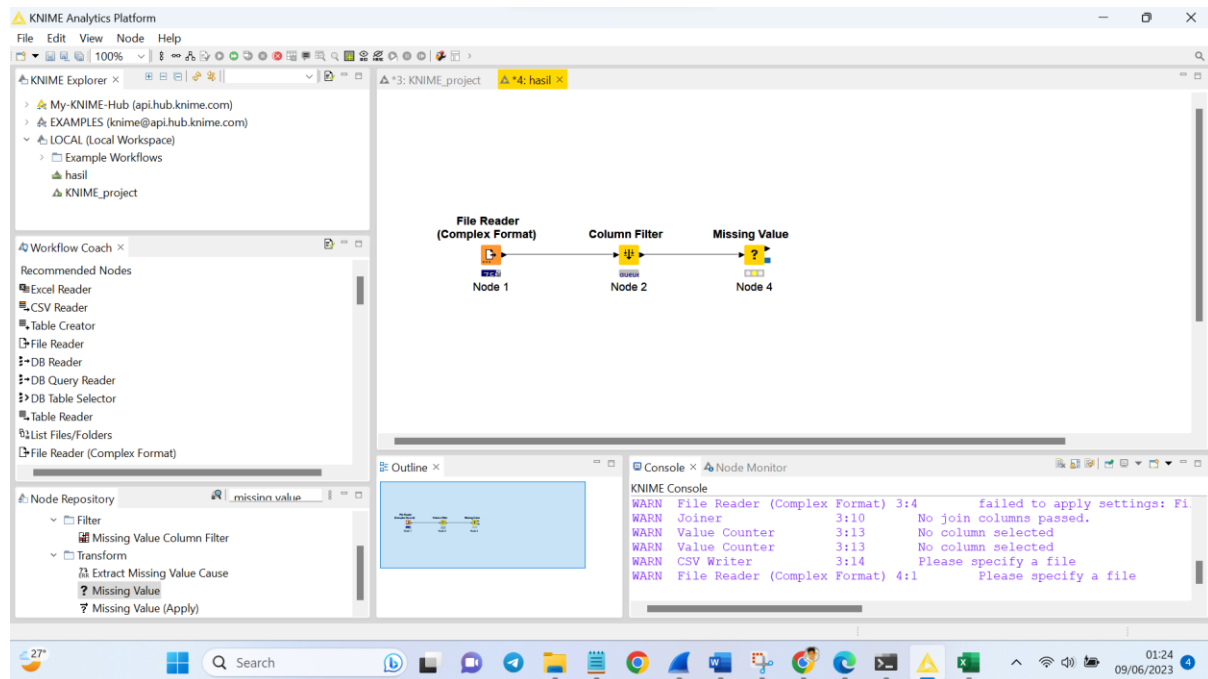
The screenshot shows an Excel spreadsheet with the following data:

No.	Time	Source	Destination	Protocol	Length	Info
1	0.000000	192.168.50.50	192.168.50.88	DNS	79	Standard query 0x5e82 A clients2.google.com"
2	0.802921	192.168.50.19	192.168.50.88	DNS	81	Standard query 0x14f4 A client-cf.dropbox.com"
3	1.848537	192.168.50.88	8.8.4.4	DNS	97	Standard query 0xfa0b A updatekeepalive.mcafee.com OPT"
4	2.096566	192.168.50.51	192.168.50.88	DNS	84	Standard query 0x4fe5 A www.google-analytics.com"
5	2.848113	192.168.50.88	192.168.50.51	DNS	86	Standard query response 0x4234 Server failure A updatekeepalive.mcafee.com"
6	Time	Source	Destination	Protocol	Length	Info
7	309.509587	192.168.50.1	192.168.50.88	TCP	74	48221 > 22 [SYN] Seq=0 Win=29200 Len=0 MSS=1460 SACK_PERM TSval=341949681 TSecr=0 WS=64"
8	309.509587	192.168.50.88	192.168.50.1	TCP	54	22 > 48221 [RST, ACK] Seq=1 Ack=1 Win=0 Len=0"
9	312.263619	192.168.50.1	192.168.50.88	TCP	74	48222 > 22 [SYN] Seq=0 Win=29200 Len=0 MSS=1460 SACK_PERM TSval=341950369 TSecr=0 WS=64"
10	312.263647	192.168.50.88	192.168.50.1	TCP	54	22 > 48222 [RST, ACK] Seq=1 Ack=1 Win=0 Len=0"
11	462.026672	192.168.50.88	142.104.64.205	TCP	74	38544 > 22 [SYN] Seq=0 Win=29200 Len=0 MSS=1460 SACK_PERM TSval=195132807 TSecr=0 WS=128"
12	463.024215	192.168.50.88	142.104.64.205	TCP	74	[TCP Retransmission] 38544 > 22 [SYN] Seq=0 Win=29200 Len=0 MSS=1460 SACK_PERM TSval=195133057 TSecr=0 WS=128"
13	472.199177	192.168.50.88	142.104.64.205	TCP	74	58842 > 1922 [SYN] Seq=0 Win=29200 Len=0 MSS=1460 SACK_PERM TSval=195135350 TSecr=0 WS=128"
14	473.196245	192.168.50.88	142.104.64.205	TCP	74	[TCP Retransmission] 58842 > 1922 [SYN] Seq=0 Win=29200 Len=0 MSS=1460 SACK_PERM TSval=195135600 TSecr=0 WS=128"
15	475.200246	192.168.50.88	142.104.64.205	TCP	74	[TCP Retransmission] 58842 > 1922 [SYN] Seq=0 Win=29200 Len=0 MSS=1460 SACK_PERM TSval=195136101 TSecr=0 WS=128"
16	479.212237	192.168.50.88	142.104.64.205	TCP	74	[TCP Retransmission] 58842 > 1922 [SYN] Seq=0 Win=29200 Len=0 MSS=1460 SACK_PERM TSval=195137104 TSecr=0 WS=128"
17	487.228249	192.168.50.88	142.104.64.205	TCP	74	[TCP Retransmission] 58842 > 1922 [SYN] Seq=0 Win=29200 Len=0 MSS=1460 SACK_PERM TSval=195139108 TSecr=0 WS=128"

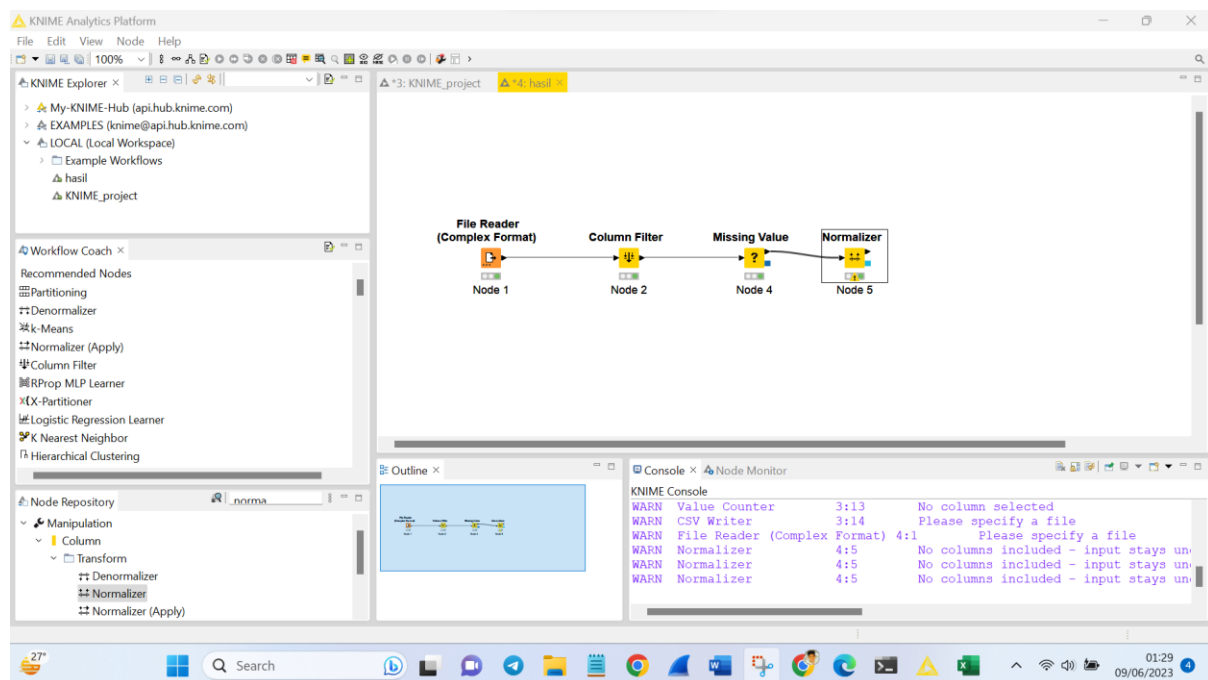


## 11. Data Pre Processing

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.

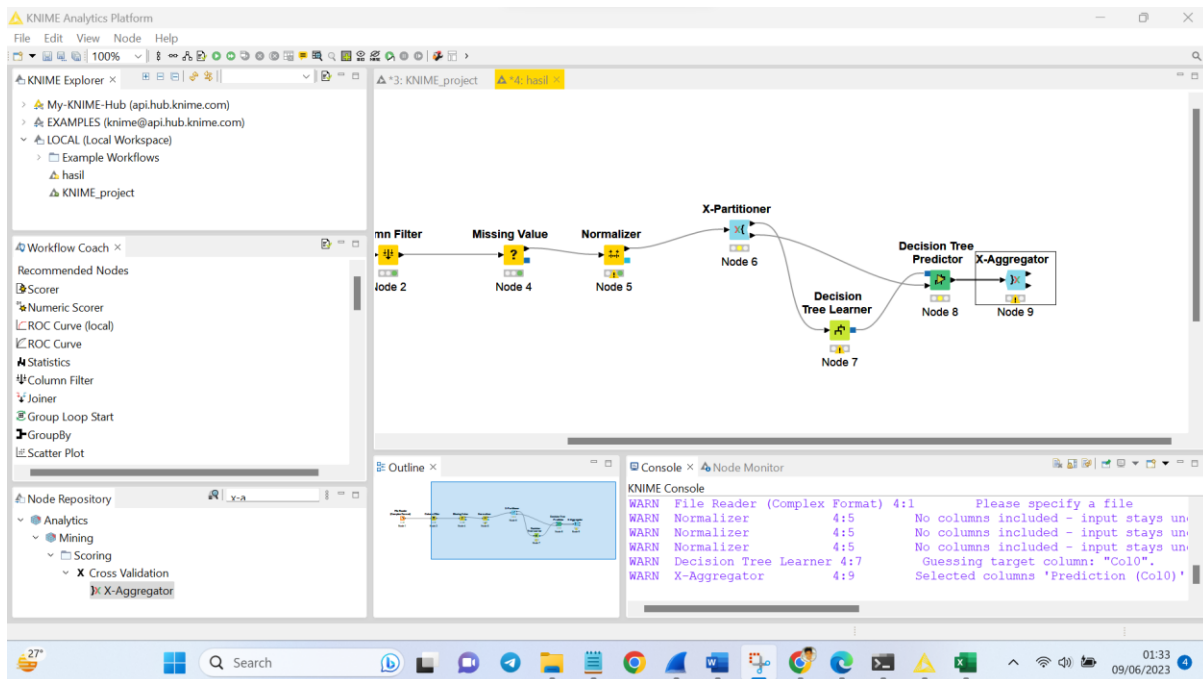


12. Proses data transformation, pada proses ini data akan diubah ke format yang sesuai untuk proses data mining. Node yang digunakan pada tahap ini yaitu Normalizer. Berikut konfigurasinya



### 13. Data Mining

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



14. Node Scorer yang didalamnya terdapat perhitungan untuk melihat seberapa baik model ini dengan menggunakan teknik confusion matrix. Berikut konfigurasiannya.

