

# **Laporan UTS Proyek B**

## **IBDA 3122 - Knowledge Discovery**



Oleh:

Moody Asyer - 191900154

**Program Studi IT and Big Data Analytics**  
**Calvin Institute of Technology**  
**Jakarta**  
**2022**

## I. Pengantar

Bursa Efek Indonesia (BEI) merupakan istilah yang terkenal pada dunia saham di Indonesia. Jumlah investor pasar modal sudah mencapai angka 7,49 juta investor per akhir Desember 2021, naik di atas 90% dibanding tahun sebelumnya (2019: 2,4 juta; 2020: 3,8 juta). Sektor perusahaan yang terdapat di BEI salah satunya adalah manufaktur. Kontribusi sektor manufaktur dalam produk domestik bruto (PDB) tercatat yang paling besar dibandingkan sektor lainnya yakni 19,7%, meskipun Indonesia tidak luput dari pandemi Covid-19. Pada perdagangan saham, harga mengalami fluktuasi naik maupun turun. Pergerakan harga saham pada dasarnya tidak dapat diprediksi secara pasti dan berfluktuasi seiring dengan bertambahnya waktu dan situasi yang berkembang.

Dalam proyek ini, diketahui bahwa grup investor XYZ ingin berinvestasi saham jangka pendek (1 bulan/30 hari) pada Unilever (UNVR), Indofood (INDF), Astra (ASII), dan Semen (SMGR). Oleh karena itu, kita akan menggunakan dataset keempat saham untuk memprediksi/melakukan perkiraan pada tiap nilai saham tersebut kedepannya.

## II. Dataset

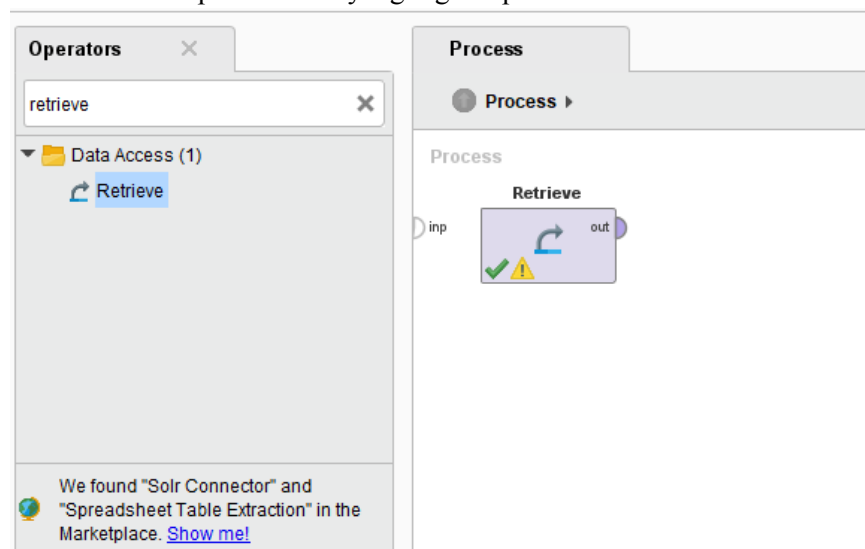
Dataset yang digunakan untuk proyek kali ini adalah dataset pergerakan harian saham UNVR, INDF, ASII, dan SMGR selama 5 tahun (sampai 22 Feb 22) dalam bentuk/format csv:

- UNVR.csv
- INDF.csv
- ASII.csv
- SMGR.csv

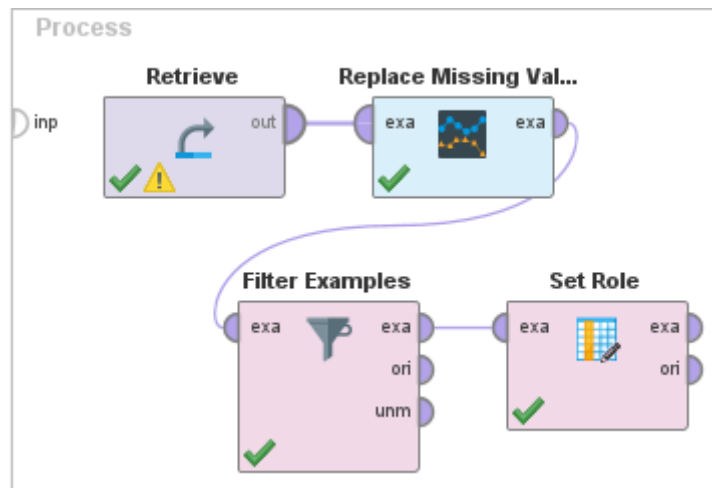
## III. Langkah Percobaan

Pada proyek ini, langkah-langkah yang perlu dilakukan untuk melakukan perkiraan pada nilai saham adalah sebagai berikut.

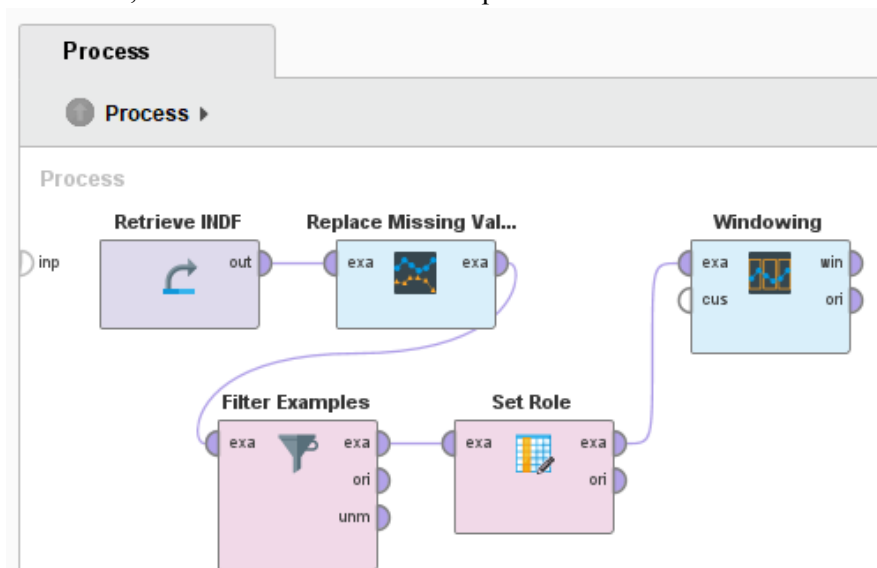
1. Hal pertama yang saya lakukan adalah membuat file csv data test dari setiap dataset yang berisi 300 hari/data terakhir
2. Lalu kita mempersiapkan dataset pergerakan harian nilai saham dengan operator 'retrieve' dan pilih dataset yang ingin diprediksi.



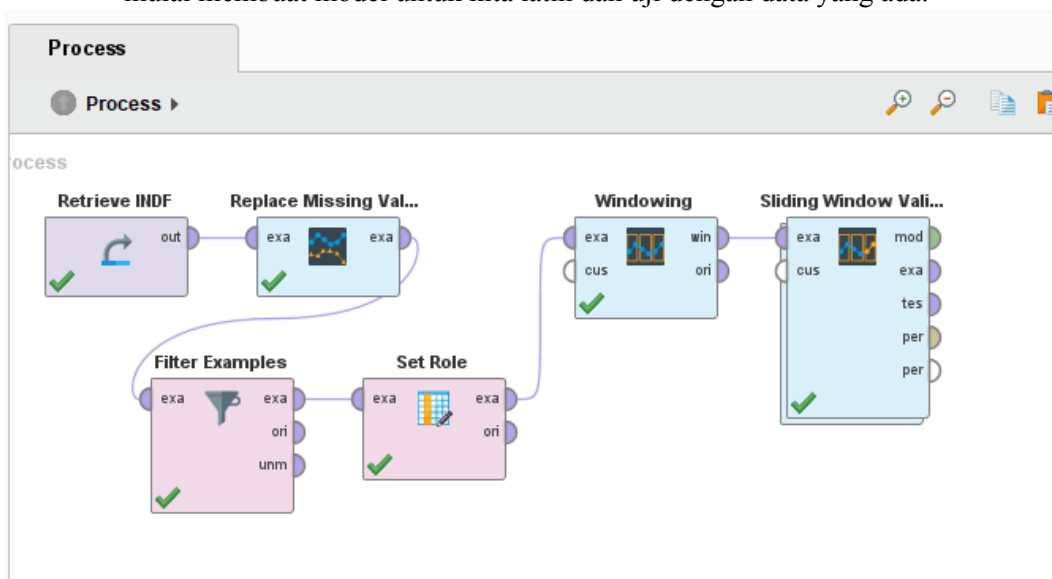
3. Selanjutnya, dilakukan pra-pemrosesan data berupa pembersihan terlebih dahulu. Dalam kasus kali ini saya mengganti missing value dengan rata-rata (tergantung kolom) dengan operator 'replace missing values', menghapus data yang memiliki nilai 0 karena data tersebut merupakan kesalahan, dan men-set role 'id' untuk date dan 'label' untuk close dengan operator 'Set Role'.



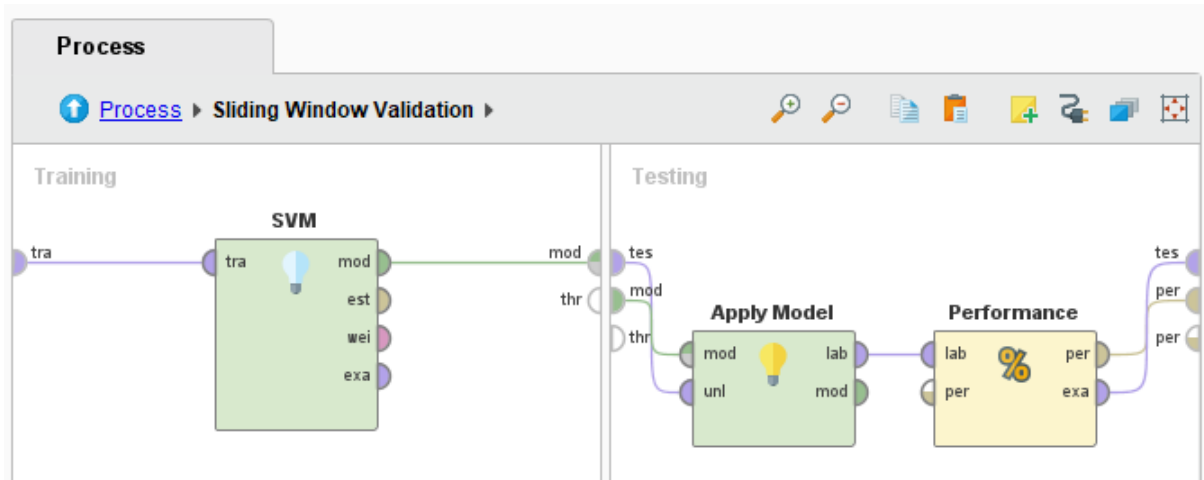
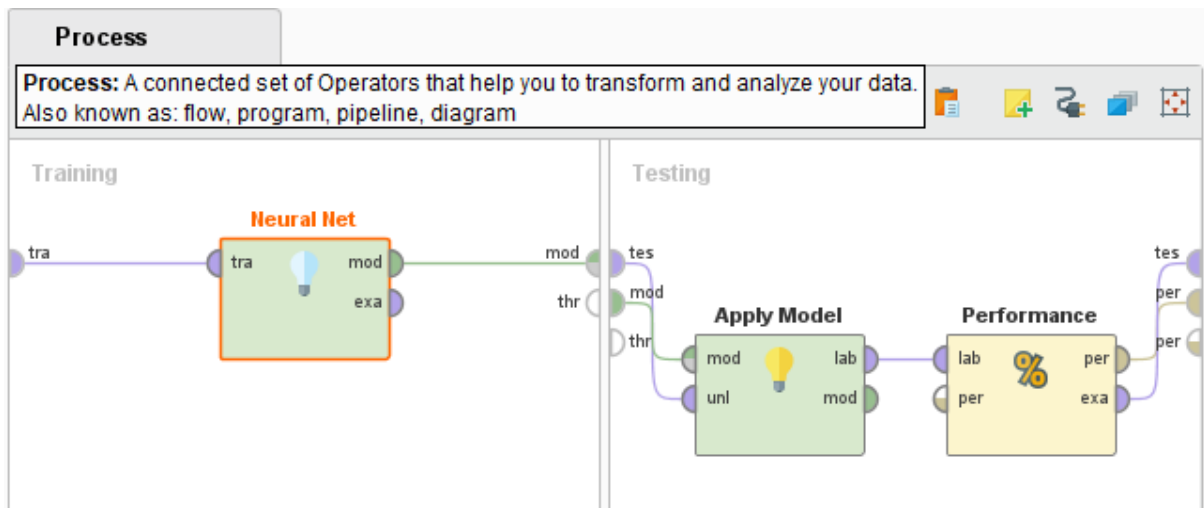
- Gunakan operator 'Windowing' untuk memproses seperti filter attribut, menentukan window, dan horizon sebelum menerapkan model.



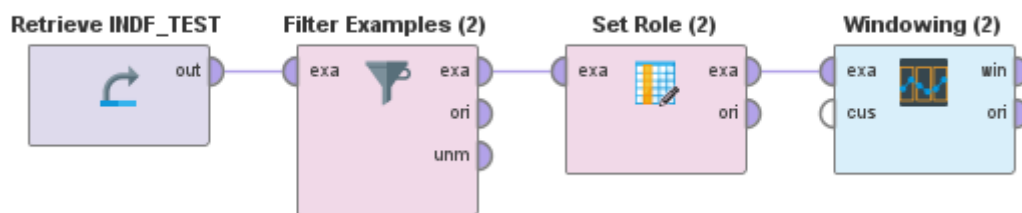
- Setelah itu dengan menggunakan operator 'Sliding Window Validation' kita akan mulai membuat model untuk kita latih dan uji dengan data yang ada.



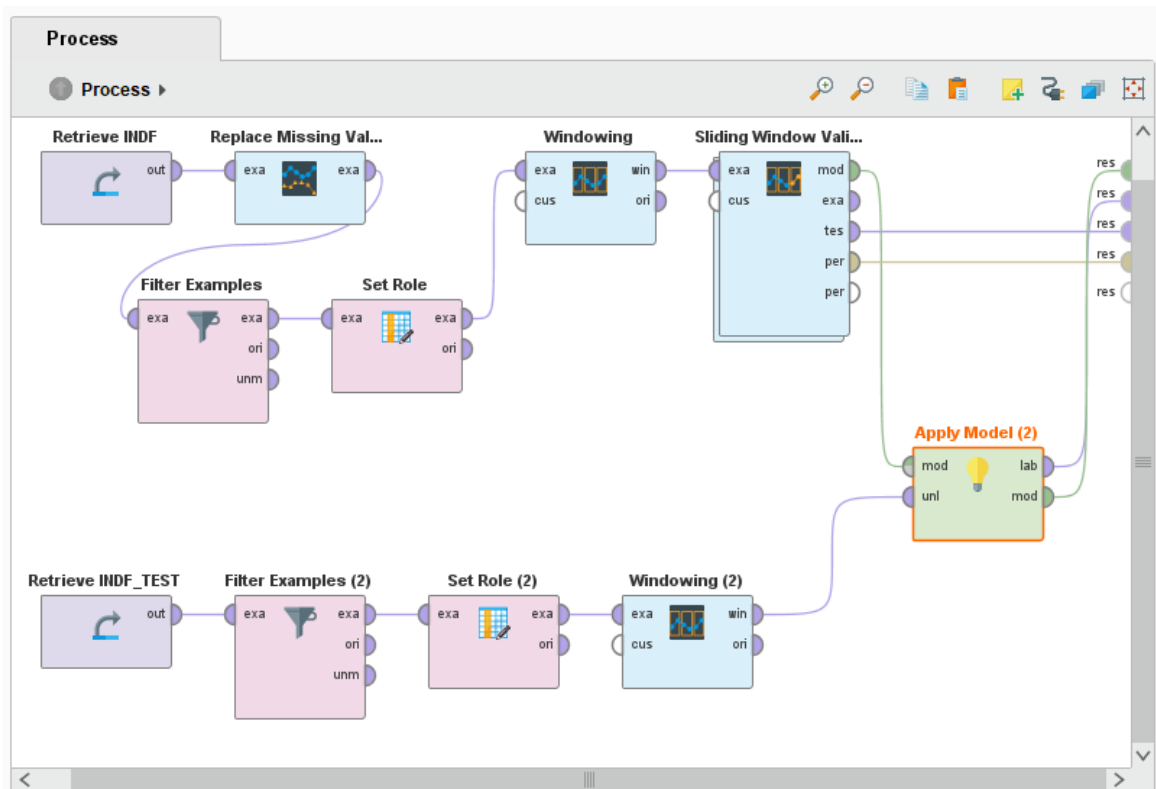
Pada kasus kali ini, kita menggunakan 2 model machine learning untuk memprediksi, yaitu Neural Network dan SVM



- Selanjutnya kita mempersiapkan data uji untuk test model kita, persiapan data test sama seperti langkah 1-3.



- Kemudian kita menghubungkan model pada operator 'Sliding Window Validation' dengan data test dari operator 'Windowing' menggunakan operator 'Apply Model'.



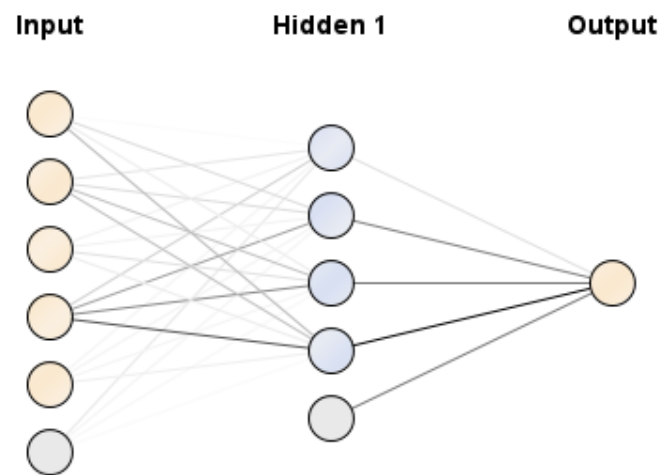
8. Jalankan model dengan memencet tombol 'run' dan kita akan mendapatkan hasil perkiraan. Untuk menemukan hasil terbaik, maka kita harus terus mencoba/mengganti parameter dari model dan menjalankan model kembali untuk melihat akurasi baru yang didapatkan hingga didapatkan akurasi yang dirasa sudah baik.

#### IV. Hasil percobaan dan Analisa

Setelah melakukan percobaan, didapatkan hasil percobaan dari keempat data saham tersebut sebagai berikut.

##### - UNVR dengan Neural Network





## prediction\_trend\_accuracy

prediction\_trend\_accuracy: 0.961 +/- 0.194 (micro average: 0.961)

### - UNVR dengan SVM



# Kernel Model

Total number of Support Vectors: 56

Bias (offset): 4292.946

$w[\text{Open} - 0] = 8.016$

$w[\text{High} - 0] = 8.439$

$w[\text{Low} - 0] = 8.197$

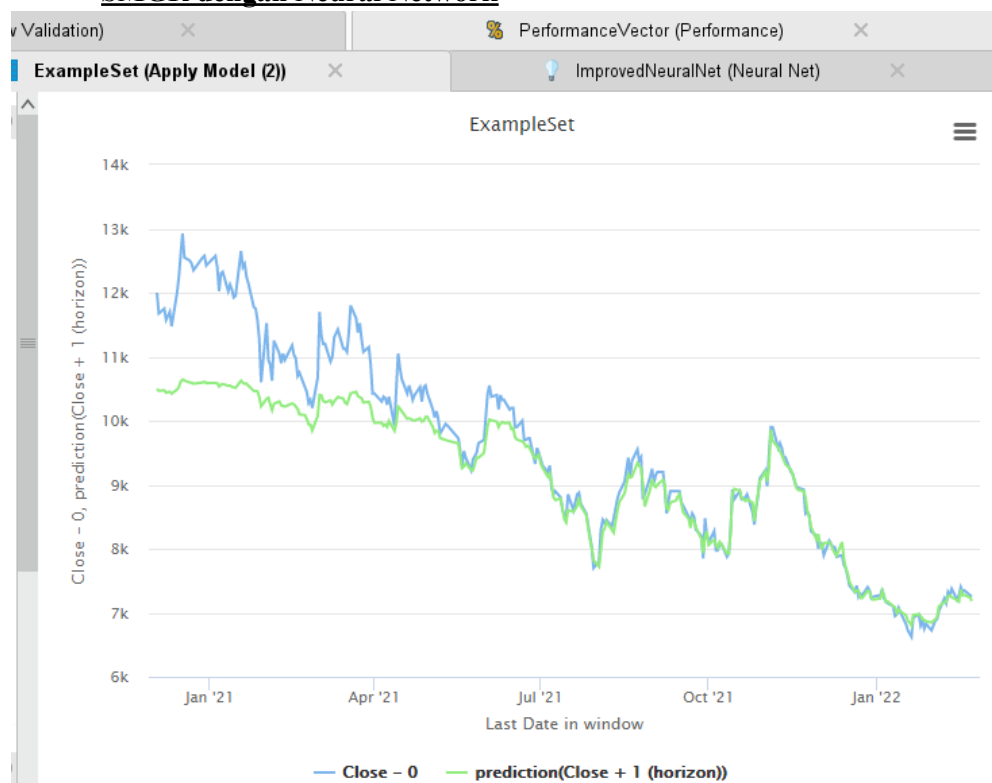
$w[\text{Adj Close} - 0] = 8.323$

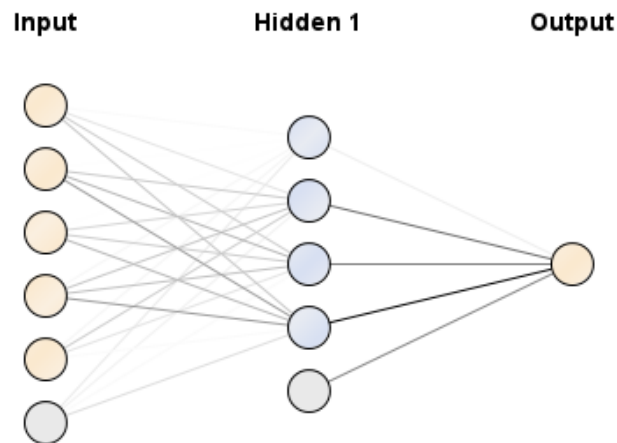
$w[\text{Volume} - 0] = 3.375$

## prediction\_trend\_accuracy

prediction\_trend\_accuracy: 0.515 +/- 0.502 (micro average: 0.515)

### - SMGR dengan Neural Network





## prediction\_trend\_accuracy

prediction\_trend\_accuracy: 0.942 +/- 0.235 (micro average: 0.942)

### - SMGR dengan SVM





# Kernel Model

Total number of Support Vectors: 52

Bias (offset): 7928.846

$w[\text{Open} - 0] = 8.596$

$w[\text{High} - 0] = 8.757$

$w[\text{Low} - 0] = 8.559$

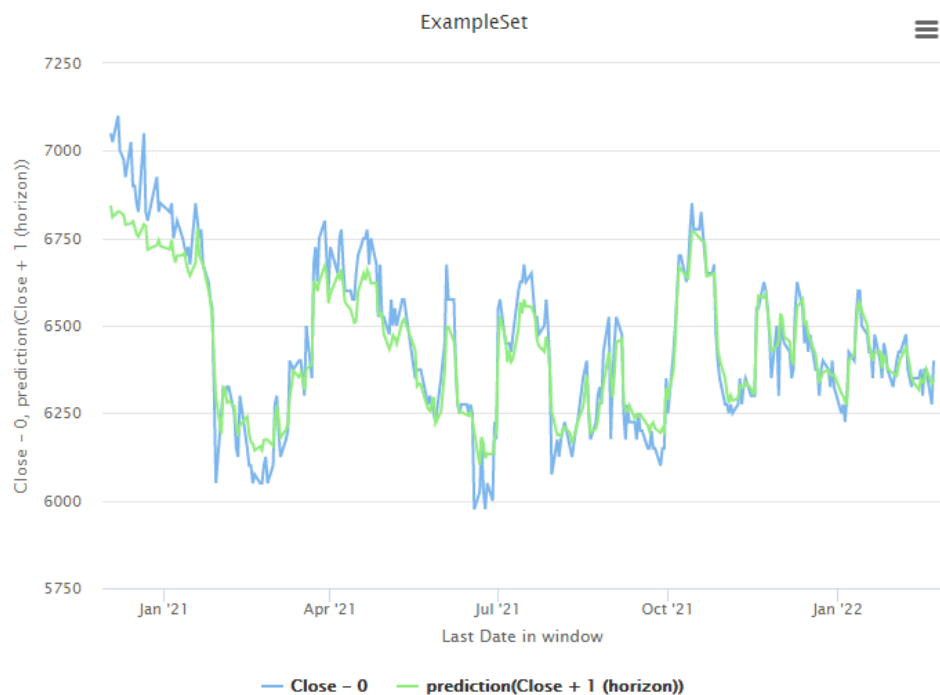
$w[\text{Adj Close} - 0] = 8.630$

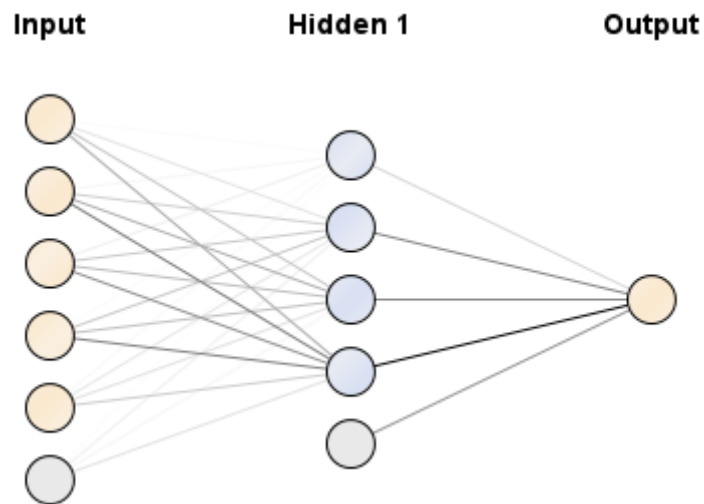
$w[\text{Volume} - 0] = 3.172$

## prediction\_trend\_accuracy

prediction\_trend\_accuracy: 0.596 +/- 0.493 (micro average: 0.596)

### - INDF dengan Neural Network





## prediction\_trend\_accuracy

prediction\_trend\_accuracy: 0.931 +/- 0.254 (micro average: 0.931)

### - INDF dengan SVM



# Kernel Model

Total number of Support Vectors: 58

Bias (offset): 6412.069

$w[\text{Open} - 0] = 6.620$

$w[\text{High} - 0] = 7.265$

$w[\text{Low} - 0] = 6.702$

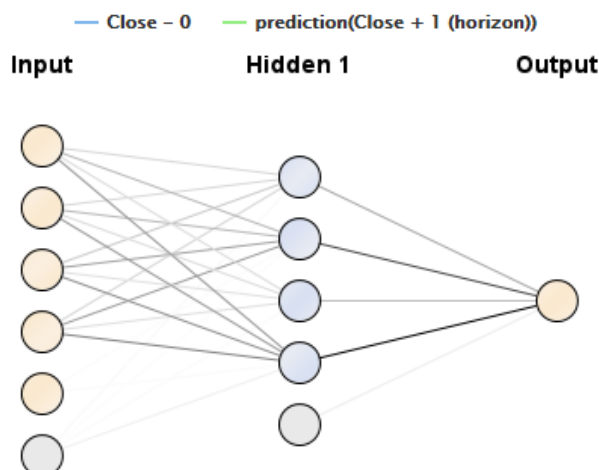
$w[\text{Adj Close} - 0] = 7.545$

$w[\text{Volume} - 0] = 2.954$

## prediction\_trend\_accuracy

prediction\_trend\_accuracy: 0.646 +/- 0.441 (micro average: 0.646)

### - ASII dengan Neural Network



## prediction\_trend\_accuracy

prediction\_trend\_accuracy: 0.942 +/- 0.234 (micro average: 0.942)

### - ASII dengan SVM



## Kernel Model

Total number of Support Vectors: 81

Bias (offset): 5537.103

w[Open - 0] = 13.177

w[High - 0] = 13.531

w[Low - 0] = 13.364

w[Adj Close - 0] = 13.610

w[Volume - 0] = 4.239

## prediction\_trend\_accuracy

prediction\_trend\_accuracy: 0.714 +/- 0.454 (micro average: 0.714)

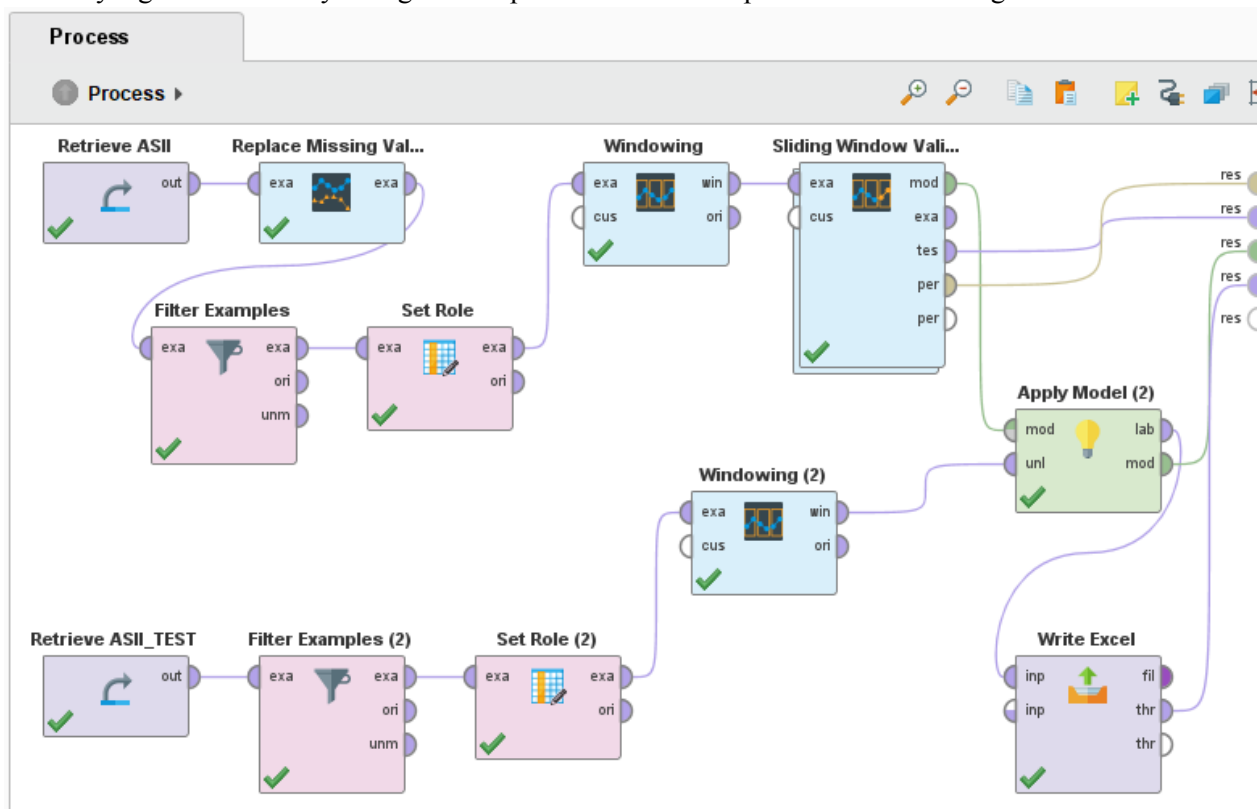
Setelah menerapkan semua model pada setiap dataset, maka ringkasan/hasil keseluruhan dari akurasi tiap model untuk tiap dataset saham adalah sebagai berikut.

	UNVR	SMGR	INDF	ASII
<b>Neural Net</b>	96.1%	94.2%	93.1%	94.2%
<b>SVM</b>	51.5%	59.6%	64.6%	71.4%

Dari percobaan yang sudah dilakukan kita dapat melihat hasil dari percobaan pada keempat dataset pergerakan harian saham bahwa dalam kasus ini model/algorithm Neural Network memiliki kemampuan yang jauh lebih baik dibandingkan dengan SVM untuk keempat jenis saham karena semua akurasi berada diatas 90%, sedangkan untuk SVM berada di akurasi kisaran 50-70%. Hal unik yang saya temukan adalah akurasi untuk saham ASII ternyata memiliki akurasi yang cukup baik menggunakan model SVM (71.4%). Hal lain yang mempengaruhi perbedaan ini adalah dari jenis saham, harganya, isi data yang berbeda (seperti ada data 0 maupun null dan missing value, dll).

## V. Evaluasi

Hal terakhir yang dilakukan juga adalah melakukan evaluasi terhadap setiap model dengan mencari MAPE (mean absolute percentage error) untuk memastikan kembali/membandingkan yang data sebenarnya dengan hasil prediksi. Maka didapat hasil MAPE sebagai berikut.



```
def mean_absolute_percentage_error(y_true, y_pred):
    y_true, y_pred = np.array(y_true), np.array(y_pred)
    return np.mean(np.abs((y_true - y_pred) / y_true)) * 100
```

	UNVR	SMGR	INDF	ASII
<b>Neural Net</b>	6.332195362112 346%	3.657268757984 3555%	0.878222882883 1201%	1.411331380140 6289%
<b>SVM</b>	16.82448659928 1855%	16.69051630587 8035%	2.440730042420 235%	5.786197189579 072%

Hasil akhir dari MAPE yang didapat adalah seperti diatas, semakin kecil MAPE yang didapat maka semakin baik karena kemungkinan error yang terjadi juga akan semakin kecil dan sedikit. Dalam kasus ini kita melihat bahwa saham INDF memiliki nilai MAPE yang terbaik dibandingkan dengan saham lainnya. Secara keseluruhan kita juga melihat bahwa MAPE model yang menggunakan neural network lebih kecil dibanding MAPE model yang menggunakan SVM.

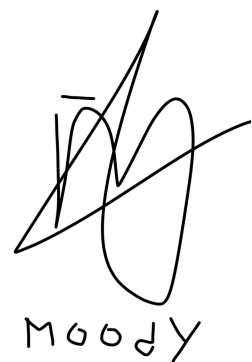
\* Untuk menguji MAPE dapat langsung menjalankan file 'MAPE Evaluation.ipynb'.

## VI. Kesimpulan

Berdasarkan percobaan yang sudah dilakukan seperti diatas, maka dapat disimpulkan bahwa:

- Parameter yang digunakan untuk memproses seperti horizon, window size, training window size, test window size, dan step size akan mempengaruhi tingkat akurasi model. Oleh karena itu perlu dilakukan percobaan terus untuk menentukan parameter terbaik yang bisa menghasilkan akurasi terbaik.
- Secara keseluruhan hasil akurasi model Neural Network lebih tinggi dibandingkan dengan svm, semakin tinggi akurasi yang didapat maka semakin baik hasilnya. Hal ini berarti model neural network lebih baik/cocok digunakan untuk memprediksi/memperkirakan hasil dari dataset saham (yang digunakan dalam kasus ini adalah UNVR, INDF, SMGR, dan ASII) dibanding SVM.
- Semakin kecil MAPE suatu model maka semakin baik performanya dalam memprediksi karena kemungkinan terdapat dan terjadinya error lebih kecil. Dalam kasus ini model Neural Network memiliki hasil yang lebih baik dibanding SVM.

“Di Hadapan TUHAN yang hidup, saya menegaskan bahwa saya tidak memberikan maupun menerima bantuan apapun-baik lisan, tulisan, maupun elektronik-di dalam ujian ini selain daripada apa yang telah diizinkan oleh pengajar, dan tidak akan menyebarkan baik soal maupun jawaban ujian kepada pihak lain.”



MOODY