summary

sobota, 16 marca 2024 19:53

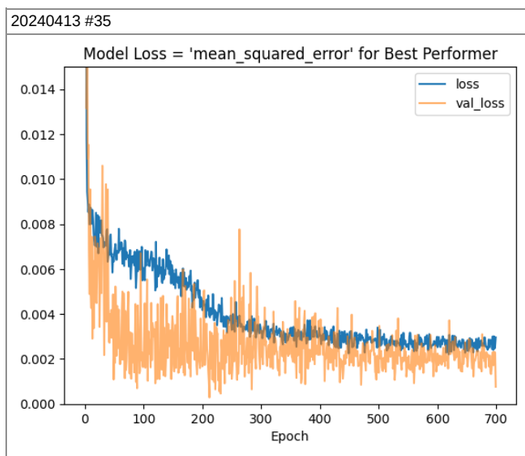
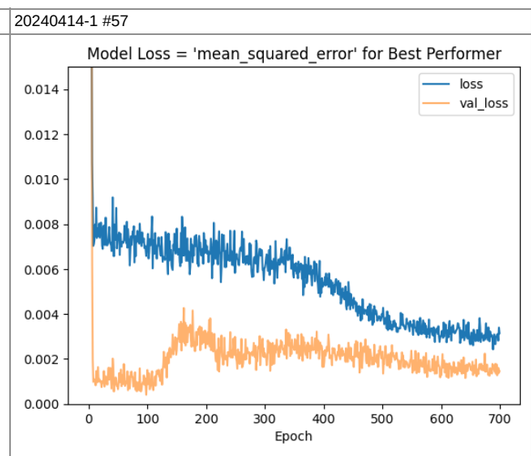
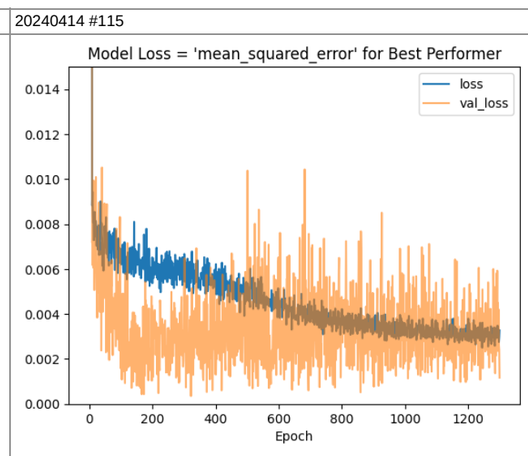
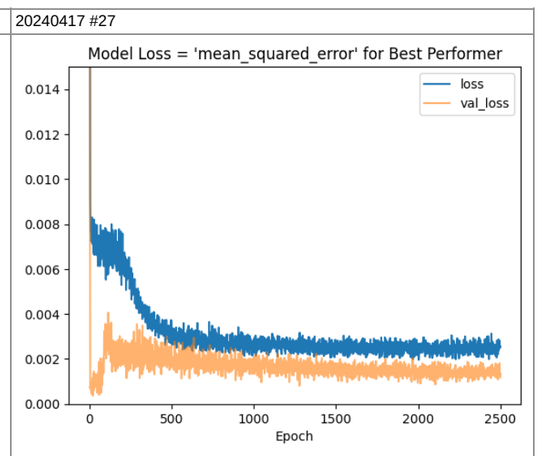
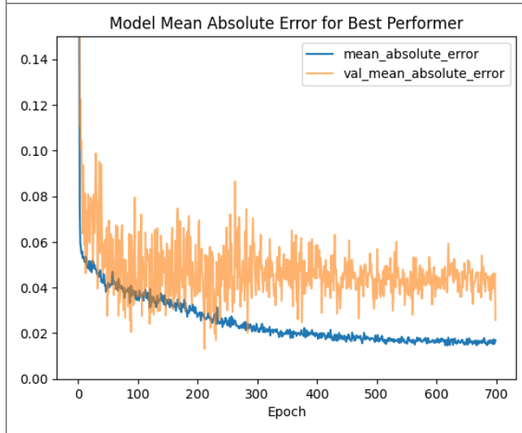
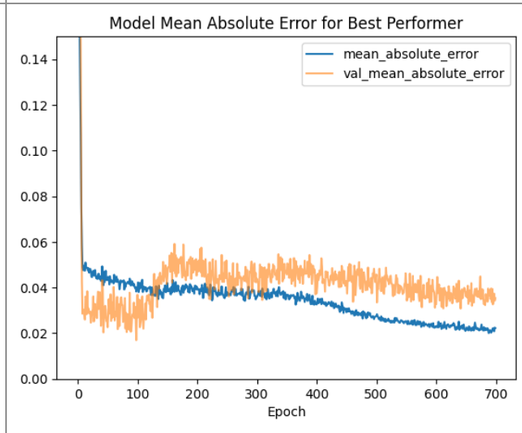
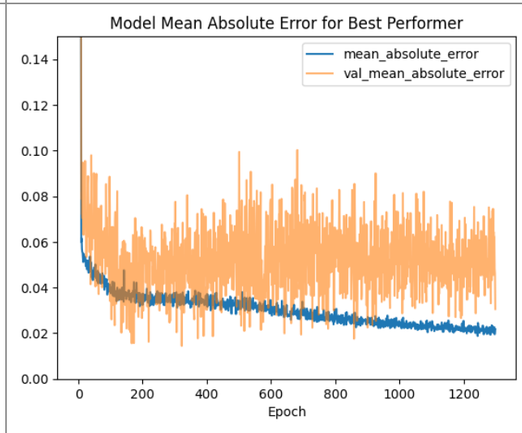
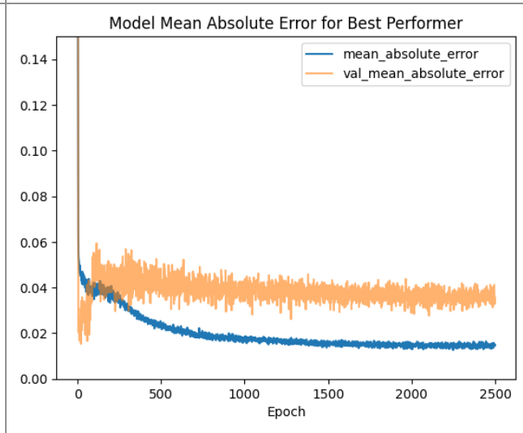
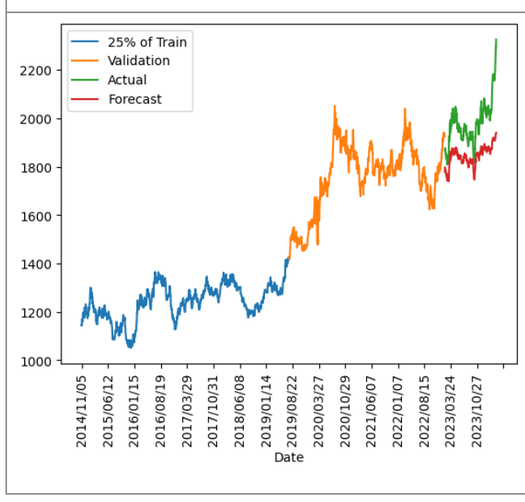
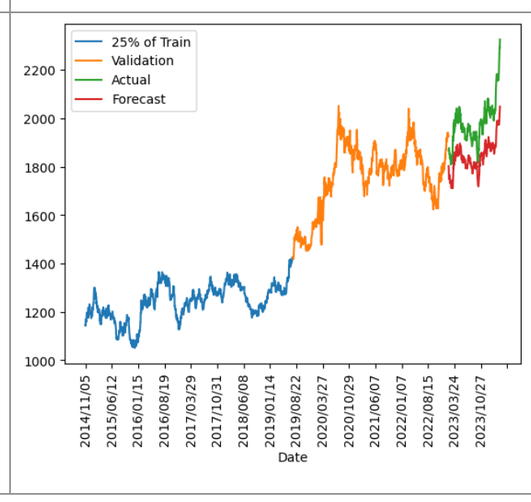
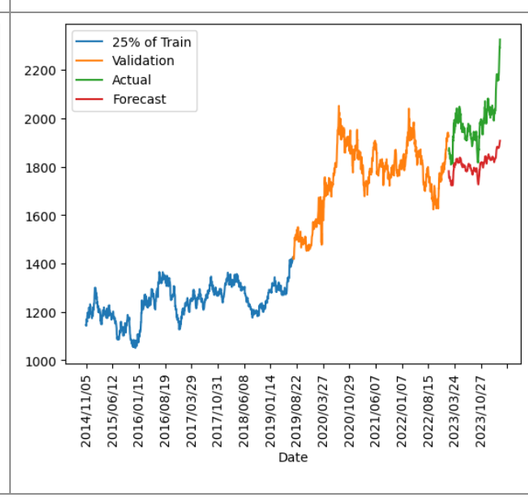
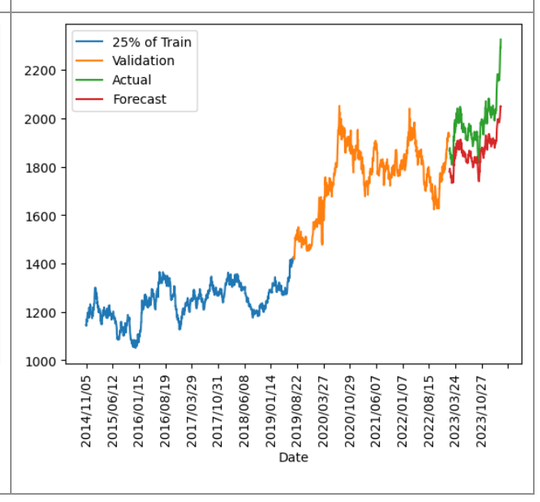
Data:
30/8/2000-05/04/2024
Train : Validation : Test -> 80:15:5

Feature building:

df['log_returns'] = np.log(1 + df['returns'])

Hyperparameters-tuning/ results summary:

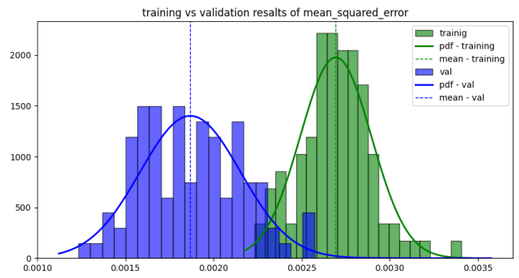
training run	2024-04-13	2024-04-14-1	2024-04-14-2	2024-04-17		Monte Carlo	2024-05-19
model # (for best performer)	35	57	115	27		N/A	N/A
lstm_nodes_1	10, **20**	10, **20**	30, **40**	40		20	20
lstm_nodes_2	**20**, 30	20, **30**	30, **40**	40		30	30
lstm_nodes_3	20, 30	20, **30**	30, **40**	40		30	30
lstm_nodes_4	N/A	N/A	**30**, 40	30, **40**		N/A	N/A
dense_nodes_1	20	20	30	**30**, 40		20	20
dropout	**0.05**, 0.1	**0.05**, 0.1	**0.05**, 0.1	0.05		0.05	0.05, 0.1, 0.2
lstm_activation	tanh, **sigmoid**	**tanh**, sigmoid	tanh, **sigmoid**	**tanh**, relu, sigmoid		tanh	tanh
dense_activation	**tanh**, sigmoid	**tanh**, relu	**tanh**, relu	**tanh**, relu, sigmoid		tanh	tanh
loss	MSE	MSE	MSE	MSE		MSE	MSE
optimizer	adam	adam	adam	adam		adam	adam
learning_rate	0.001	0.0001	0.0005	0.0001		0.00005	0.001
batch_size	32	64	64	64		64	64
stopped_epoch (for best performer)	699	699	1299	2499		N/A	N/A
early_stopping_monitor	val_loss	val_loss	val_loss	val_loss		val_loss	val_loss
patience	200	200	300	500		500	500
start_from_epoch	500	500	1000	2000		1500	1500
Train Score [RMSE] (for best performer)	$60.60	$68.57	$54.02	$40.16		N/A	N/A
Validation Score [RMSE] (for best performer)	$77.01	$99.92	$93.67	$79.57		N/A	N/A
Test Score [RMSE] (for best performer)	$144.70	$141.50	$179.56	$120.13		N/A	N/A
# of Monte Carlo Replications	N/A	N/A	N/A	N/A		100	N/A

20240413 #35	20240414-1 #57	20240414 #115	20240417 #27
			
			
			

Monte Carlo simulation results:

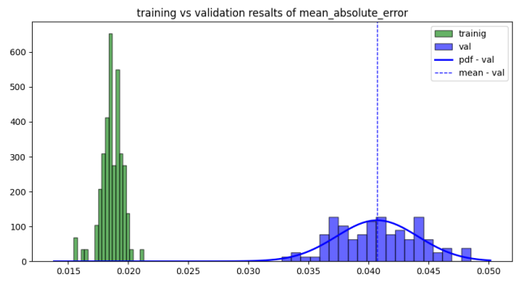
MSE

p_value threshold: 0.05
(not applicable to Anderson test)
p_value for Anderson test: 0.05
---
p_value for mean_squared_error: 0.1386
mean_squared_error is normally distributed based on normaltest
(Shapiro-Wilk test is based on the correlation ratio)
---
mean: 0.0027 standard deviation: 0.0002
---
p_value threshold: 0.05
(not applicable to Anderson test)
p_value for Anderson test: 0.05
---
p_value for val_mean_squared_error: 0.2839
val_mean_squared_error is normally distributed based on normaltest
(Shapiro-Wilk test is based on the correlation ratio)
---
mean: 0.0019 standard deviation: 0.0003

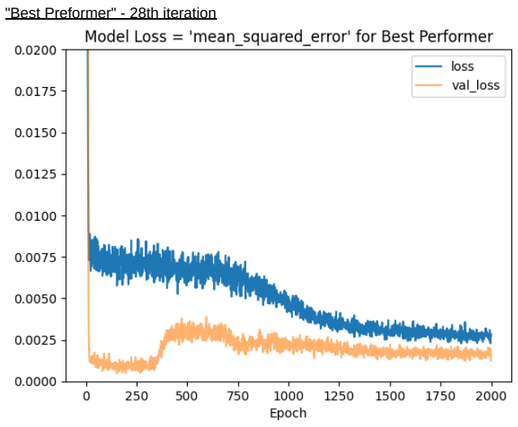


MAE

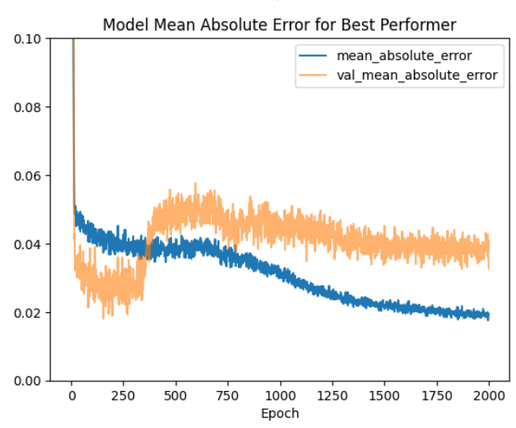
p_value threshold: 0.05
(not applicable to Anderson test)
p_value for Anderson test: 0.05
---
p_value for mean_absolute_error: 0.0008
mean_absolute_error is not normally distributed based on shapiro-wilk
(Shapiro-Wilk test is based on the correlation ratio)
---
p_value threshold: 0.05
(not applicable to Anderson test)
p_value for Anderson test: 0.05
---
p_value for val_mean_absolute_error: 0.5363
val_mean_absolute_error is normally distributed based on normaltest
(Shapiro-Wilk test is based on the correlation ratio)
---
mean: 0.0407 standard deviation: 0.0034



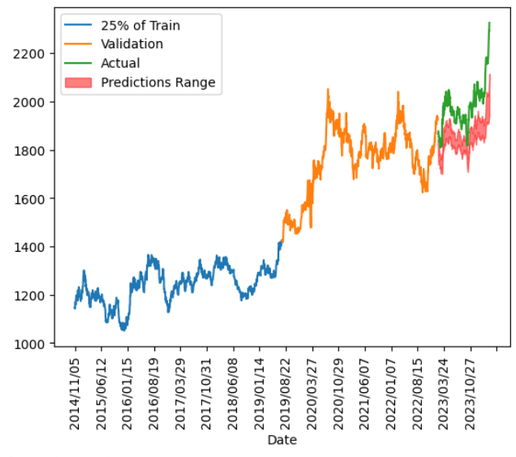
"Best Preformer" - 28th iteration



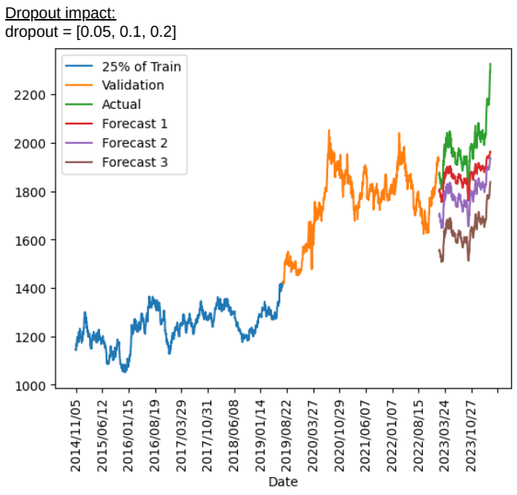
Model Mean Absolute Error for Best Performer



Monte Carlo Predictions Range



Dropout impact:
dropout = [0.05, 0.1, 0.2]



RMSE Score	Model 1	Model 2	Model 3
Train	$33.54	$84.68	$162.18
Validation	$62.15	$159.64	$295.74
Test	$124.59	$200.16	$342.98

Monte Carlo Simulation
A method for solving problems using random sampling. It repeatedly generates random inputs to a model to calculate a range of possible outcomes, helping to approximate complex systems or solve mathematical problems where deterministic methods are not feasible.