run	model	epochs	batch size	hidden size	loss	optimizer	lr	train loss	test loss
1	LSTM, num_layers=2, dropout=0.3, bidirectional=True	10	32	128	MSE, reduction='mean'	adam, lr=0.0001	step_size=200, gamma=0.1	0.2224	0.4682
2	LSTM, num_layers=4, dropout=0.5, bidirectional=True	15	32	256	MSE, reduction='mean'	adam, lr=0.0002	step_size=200, gamma=0.01	0.1834	0.3798
3	LSTM, num_layers=4, dropout=0.1, bidirectional=True	50	16	128	MSE, reduction='mean'	adam, lr=0.001	step_size=300, gamma=0.2	0.1353	**0.3702**
4	LSTM, num_layers=3, dropout=0.7, bidirectional=False	10	64	128	MSE, reduction='mean'	adam, lr=0.0005	step_size=100, gamma=0.1	0.2053	0.4048
5	LSTM, num_layers=3, dropout=0.5, bidirectional=True	25	32	256	MSE, reduction='mean'	adam, lr=0.001	step_size=300, gamma=0.2	0.1603	0.3823
6	LSTM, num_layers=2, dropout=0.5, bidirectional=True	10	16	128	MSE, reduction='mean'	adam, lr=0.001	step_size=300, gamma=0.2	0.196	0.3863
7	LSTM, num_layers=2, dropout=0.4, bidirectional=True	20	32	256	MSE, reduction='mean'	adam, lr=0.0002	step_size=200, gamma=0.01	0.1668	0.3618
8	LSTM, num_layers=3, dropout=0.7, bidirectional=True	10	128	64	MSE, reduction='mean'	adam, lr=0.0001	step_size=50, gamma=0.01	0.2552	0.3938
9	Transformer, dropout = 0.5	10	128	6	MSE, reduction='mean'	adam, lr=0.0001	step_size=200, gamma=0.1	0.7458	0.6386
10	Transformer, dropout = 0.5	20	32	6	MSE, reduction='mean'	adam, lr=0.0001	step_size=200, gamma=0.01	0.448	0.504
11	Transformer, dropout = 0.5	100	32	6	MSE, reduction='mean'	adam, lr=0.0001	step_size=200, gamma=0.01	0.448	0.502
12	Transformer, dropout = 0.3	20	16	6	MSE, reduction='mean'	adam, lr=0.001	step_size=300, gamma=0.2	0.389	0.4496
13	Transformer, dropout = 0.2	15	16	6	MSE, reduction='mean'	adamW, lr=0.001, wd=0.01	step_size=200, gamma=0.1	0.389	0.4464
14	Transformer, dropout = 0.7	10	8	6	MSE, reduction='mean'	adam, lr=0.0005	step_size=400, gamma=0.3	0.3181	0.378
15	Transformer, dropout = 0.6	10	64	6	MSE, reduction='mean'	adam, lr=0.0002	step_size=100, gamma=0.05	0.503	0.5661
16	Transformer, dropout = 0.7	15	64	6	MSE, reduction='mean'	adamW, lr=0.0005, wd=0.01	step_size=400, gamma=0.3	0.6105	0.5391
17	Transformer, dropout = 0.7	15	4	6	MSE, reduction='mean'	adamW, lr=0.0005, wd=0.01	step_size=400, gamma=0.3	0.2543	**0.3175**
18	LSTM, num_layers=3, dropout=0.5, bidirectional=True	10	4	128	MSE, reduction='mean'	adam, lr=0.001	step_size=300, gamma=0.2	0.1991	0.4261