Table 1

				Loss, Accuracy, F1 and Kappa are in the format: [Training] | [Val]				
Model/Notebook ID	Model Type	Target - Days	Data	Loss	Acc	F1	Kappa	Notes
1000	LSTM	10	40LOB 5/7 days	71|90	71|63	68|63	50|44	Base Case LSTM Model 100 units with l2 regularization and dropout (40%). Learning rate at 0.001. 100 epochs. 500 windows per batch. Each window looks at 20 LOB snapshots.
1001	LSTM	10	40LOB 5/7 days	72.12| 1.00	70.67| 58.7	68|59	49|38	Reduced learning rate. Performance degraded on Val set.
1000Test	**LSTM**	**10**	**40LOB 7/9**	**70|89.65**	**71.4|64**	**67|64**	**50|46**	**Test version of Base case LSTM model. Very similar to training run.**
2000	Conv2d	5	40LOB - 5/7 days	95/107				Slow learning. Val divergence. No regularization or holdout
2010	Conv2d	10	40LOB - 5/7 days	95/118				Slow learning. Val divergence. Self generated images show model preference for mean reversion.
2011	Conv2d	10	40LOB - 5/7 days	75|108	65|52	62|52	40|26	Big improvement from turning off window standardization. Val divergence after epoch 25. Minimum loss on val was below 100. Plot of highest up/down windows good for volume and price
2012	Conv2d	10	40LOB - 5/7 days	89/106	55|46	56|47	29|18	Added dropout of 50 at last level. Val and training results closer but poorer overall.
2013	Conv2d	10	40LOB - 5/7 days	74|93	65|57	65|57	44|35	Increased Dense layer to 100. Dropped 1 max pool layer. Val diverges after epoch 45. Good graphic results.
2014	Conv2d	10	40LOB - 5/7 days	92|99	51|45	55|49	21|17	Add l2 regularization to dense layer —- network couldn't learn. Removed and increased dropout to 60. Results were worse but val tracked training better and continued to decline
2015	Conv2d	10	40LOB - 5/7 days	74/99	65|57	46|34	46|34	Increased number of epochs to 100 with the higher dropout level (60). Val diverged in range of 60-70. Better to go back to dropout of 50
2016	**Conv2d**	**10**	**40LOB - 5/7 days**	**86|88**	**61|60**	**63|60**	**41|39**	**Add back l2 regularization and shrink batch size to 50. Dropout restored to 50. 100 epochs. Good graphic results. Best overall results for Convent but still weaker than LSTM.**
2017	Conv2d	10	40LOB -5/7 days					increase learning rate to 0.01 — doesn't learn. Interrupt and Add batch normalization after dense layer with decrease in lr to 0.005 - val loss is high at first but comes down. But learning is slow and seems to be stuck - interrupt at epoch 20. Try one last time with BN and original learning rate of 0.001. Erratic and high Val loss (1.5). Interrupt after epoch 25.
2016Test	**Conv2d**	**10**	**40 LOB 7/9**	**88|87**	**59.8|61.6**	**62|62**	**41|42**	**Test Run on Model 2016 - model generalized well from Training run. Validation loss followed sawtooth pattern but declined to track training loss. Good graphic results**
2016Hybrid	Conv2d	10	20 LOB + 20 handcrafted 5/7	73.6|81	72|69	72|69	57|52	Model 2016 using handcrafted features. Top 5 levels (rather than 10) of order book with derivatives (cols 86-105) for these levels. Significantly improved results but larger gap between test and val
2016HybridTest	**Conv2d**	**10**	**20 LOB +20 handcrafted 79**	**72.7| 74.9**	**72.4|71.5**	**73|73**	**58|51**	**Test run on Model 2016Hybrid. Results generalize well. Use of the handcrafted features results in improvement across all metrics.**
3016Dilation	Conv-1D	10	40 LOB 5/7	43|42.7	86|86	86|86	80|79	Dilated convents with causal padding (simplified Wavenet architecture). Best results of any architecture tried.
3016DilationTest	**Conv1D**	**10**	**40 IOB 7/9**	**41.6| 34.05**	**85.5| 89.06**	**84|83**	**89|89**	**Test run using dilated convent**
4016	Conv-LSTM	10	40 LOB 5/7	72|86	69|62.5	68|63	49|43	First run with 50 windows per batch is learning well but runs too slowly (about 4 minutes per epoch. Interrupt at epoch 26 with loss 82.4|88.8 which is already comparable to final results of model 2016. Increase Batch size to 500 which reduces epoch time to 30seconds. At epoch 26 loss is 84.4|89.2 which is pretty close. Final results are good but training/val difference is large - more regularization needed.
4017	Conv-LSTM	10	40LOB - 5/7 days	76|84.6	66.6|62.4	66|62	47|43	Increased dropout to 60 - small improvement. Not worth pursuing further
4018	Conv-LSTM	10	40 LOB 5/7	71.3| 81.5	71|66.5	67|67	49|50	Higher learning rate 0.01(vs 0.001) leads to no learning. Smaller learning rate 0.0005 (vs 0.001) leads to improvement

4019	Conv-Lstm	10	40 LOB 5/7	72.8|81.8	68.8|64	68|64	50|46	Using the lower learning rate (0.0005), also reduce batch size from 500 to 250. BEST IMAGES FOR POSTER . Reduced batch sized doesn't seem to help when there are LSTMs. Balance between large batch for LSTM efficiency and smaller batch for avoiding getting caught in local minimums.
5002	Transfer Learning using LSTM 1000	10	40 LOB 5/7	see table below				Fitting model on 5 stocks and predicting on 1 generally worked better than fitting model on just 1 stock and predicting on that stock
5002Test	Transfer learning using LSTM 1000	10	40 lOB 7/9	see table below				
	<u>Transfer Learning</u>							
		Test 7/9						
		5->1		1->1				
		F1	K	F1	K			
	1	61|59	31|23	63|60	27|19			
	2	75|65	56|43	77|59	60|31			
	3	75|66	55|48	77|55	63.6|31.6			
	4	70|66	55|44	76|65	64.58|46.2			
	5	60|58	40|37	53|50	0|0			
		Train 5/7						
		5->1		1->1				
		F1	K	F1	K			
	1	61|62	30|17	64|61	28|23			
	2	75|68	55|46	79|61	61|27			
	3	77|68	60|50	74|58	72|28.6			
	4	71|63	55|45	77|60	65|40			
	5	62|56	42|35	55|50	0|0			
	<u>Notes:</u>							
	40 LOB - 10 levels where each level contains bid price, bid volume, ask price and ask volume							
	5/7 — train on days 0-5 and validate on days 6-7							
	7/9 — train on days 0-7 and validate on days 8-9							
	All models run on EC2 P3.Xlarge instance.							

Table 1

		Loss	Accuracy	F1	Kappa	
	LSTM	0.71|0.90	0.71|0.64	0.67|0.64	0.50|0.46	
	CNN	0.88|0.88	0.60|0.62	0.62|0.62	0.41|0.42	
	CNN-LSTM	0.72|0.81	0.69|0.64	0.67|0.64	0.50|0.46	
	Dilated-CNN	0.42|0.34	0.85|0.89	0.89|0.89	0.83|0.84	

Table 1

LSTM	100 LSTM Units -> Dropout -> Softmax			
CNN	3 Conv layers (16 filters) -> MP -> Conv Layers (32 filters) -> MP -> 100 unit FC->Softmax			
CNN-LSTM	Same as CNN model with 100 unit LSTM layer replacing FC layer			
Dilated-CNN	5 dilated Conv layers -> 100 unit FC -> Dropout ->Softmax			

Table 1
