Convolutional Neural Network Model

Report

An exploration and analysis of CNN models deployed on time series data.

By Aubrey Waddle, Data Engineer

Initial Descriptions

UVXY time series data going back to 2011. The data consists of the daily closing price of the ProShares Ultra VIX Short-Term Futures ETF (UVXY). The choice to use returns instead of raw prices is appropriate for financial time series, as it focuses on relative changes and reduces the risk of trends overwhelming the model.

Model Architecture

The time series data is scaled using MinMaxScaler, which helps the model converge more effectively by normalizing the input values to a range between 0 and 1. A sequence length of 60 is used, allowing the model to capture recent trends and dependencies in the data. The CNN consists of two convolutional layers followed by a dense layer and dropout layers for regularization. This architecture allows for effective feature extraction from the input time series while minimizing overfitting. The model is trained with 20% of the data reserved for validation, enabling an evaluation of the model's generalization performance. The model forecasts 3294 steps ahead, which corresponds to the number of rows in the dataset. This ambitious forecasting horizon highlights the need to evaluate long-term trends and forecast stability.

Model Tuning

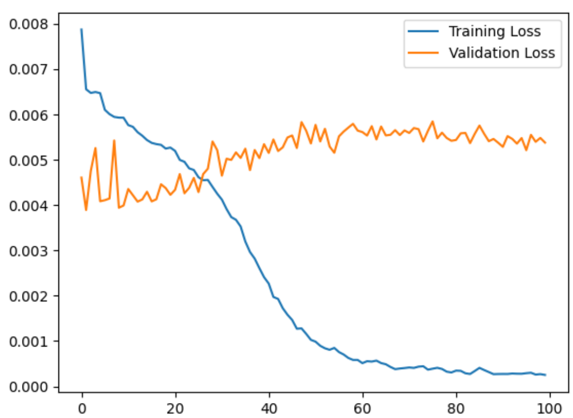
Model Rendition 1.0 Hyperparameters

Sequence Length	60
Conv1D Filters	128
Dropout	0.3
Conv1D Filters	64
Dense	100
Epochs	100
batch_size	16
validation_split	0.2

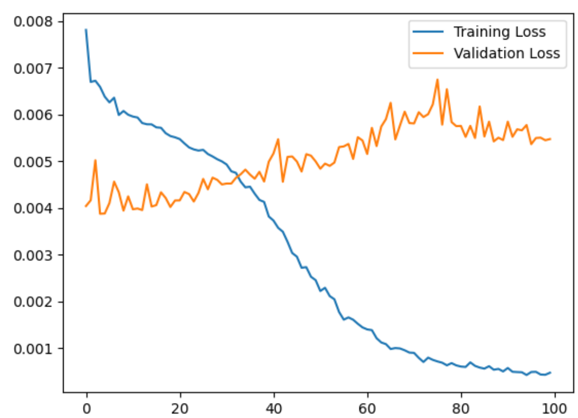
Model Rendition 1.1 Hyperparameters

Sequence Length	60
Conv1D Filters	64
Dropout	0.3
Conv1D Filters	32
Dense	50
Epochs	100
batch_size	16
validation_split	0.2

Model Rendition 1.0 Loss Curve

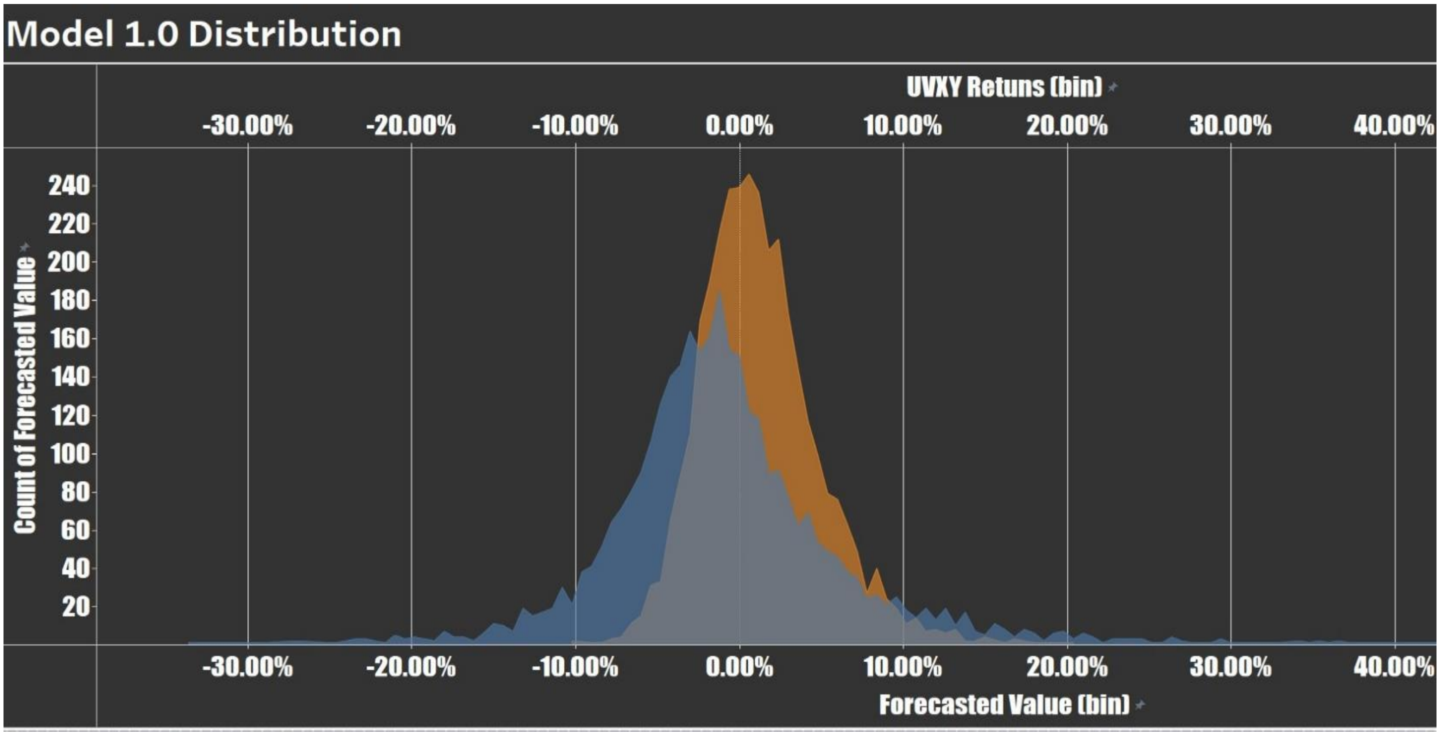


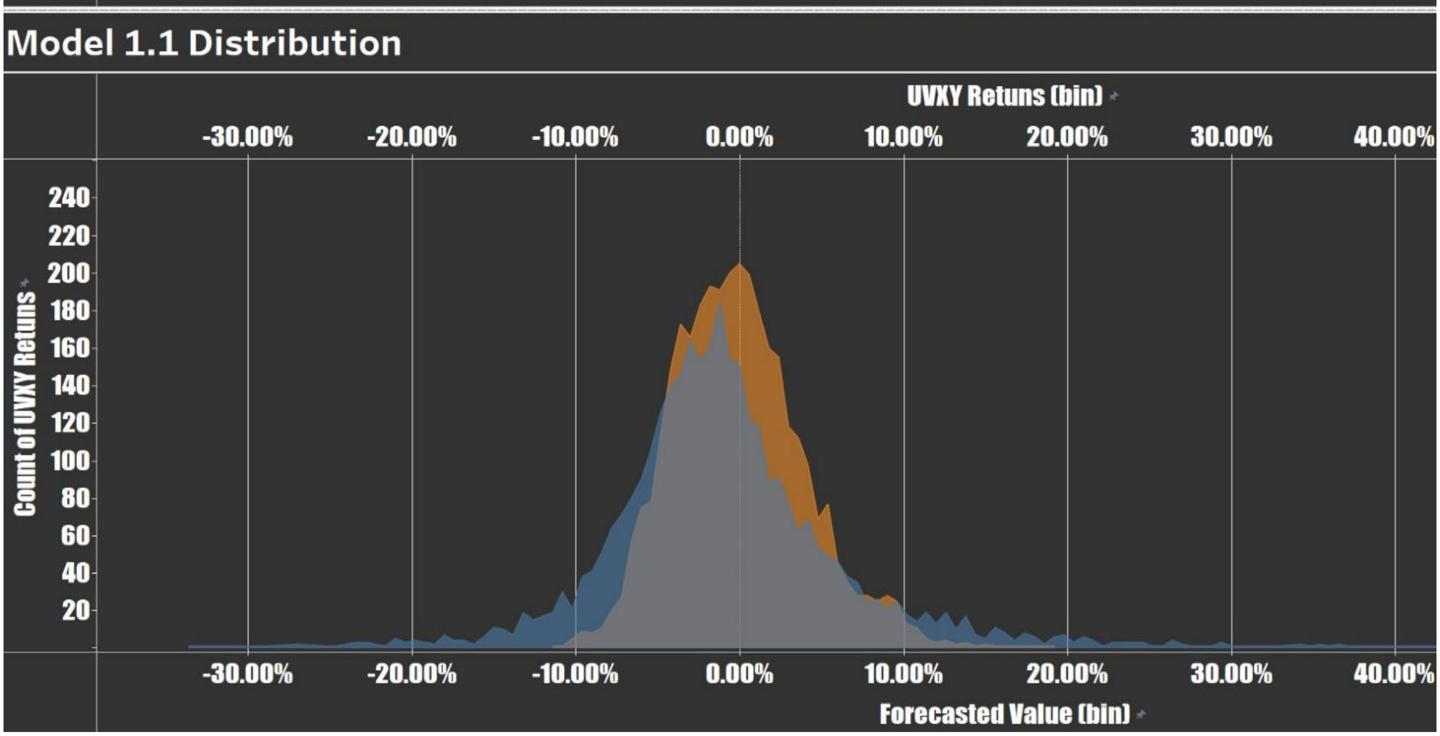
Model Rendition 1.1 Loss Curve



The **training loss** decreases steadily over the epochs, which indicates that the model is effectively learning patterns in the training data. The **validation loss** stabilizes after a certain point, but there is a noticeable gap between the training and validation losses, suggesting some level of overfitting.

The orange distribution is the forecasted data output by the model, the blue is the input data. There are 3294 rows of each.





Model 1.0 Notes

Model 1.0 Loss Curve

- **Training Loss**:
  - Decreases steadily throughout the training process, indicating that the model is learning from the training data.
- **Validation Loss**:
  - Shows a noticeable gap from the training loss after a certain number of epochs.
  - Validation loss stabilizes early but exhibits fluctuations, which may indicate overfitting as the model memorizes the training data but struggles to generalize well to unseen data.

Model 1.0 Distribution

- **Forecasted Values (Orange)**:
  - The forecasted values have a relatively narrow distribution.
  - Forecasts are concentrated around a specific range, indicating the model's predictions lack variability.
- **Input Data (Blue)**:
  - The input data shows a much wider distribution, reflecting the inherent volatility and variability of the UVXY returns.

Model 1.0

- **Strengths**:
  - Consistently decreasing training loss indicates effective learning from the training data.
- **Weaknesses**:
  - Larger validation loss and fluctuations suggest poor generalization and overfitting.
  - Narrow forecast distribution indicates the model struggles to capture the volatility and patterns in the UVXY returns data.

Model 1.1 Notes

Model 1.1 Loss Curve

- Training Loss:
  - Similar to Model 1.0, it decreases consistently over epochs, reflecting learning.
- Validation Loss:
  - Stabilizes earlier and fluctuates less compared to Model 1.0, suggesting better generalization.
  - The gap between training and validation loss is smaller in Model 1.1, indicating that overfitting has been reduced.

Model 1.1 Distribution

- **Forecasted Values (Orange)**:
  - Still shows a narrower distribution than the input data but has a slightly wider spread compared to Model 1.0.
  - Predictions begin to capture more of the variability in the data.
- **Input Data (Blue)**:
  - Similar to Model 1.0, the input data maintains a wide distribution, emphasizing the model's challenge in fully replicating the volatility.

Model 1.1

- **Strengths**:
  - Improved validation loss stability and a smaller gap between training and validation losses suggest reduced overfitting.
  - Forecasted values show better variability, reflecting an improved ability to capture patterns in the data.
- **Weaknesses**:
  - The forecast distribution is still narrower than the input data, indicating further room for improvement in capturing variability.

Next Steps

1. Increase Regularization:
   - Additional dropout layers or L2 regularization could further improve Model 1.1's ability to generalize without losing the ability to learn.
2. Incorporate Volatility Features:
   - Adding features like rolling standard deviations or volatility indices could help the model better represent variability in the data.
3. Longer Sequence Length:
   - Experimenting with a longer sequence length (e.g., 120) may allow the model to capture more of the temporal dependencies in the data.