EE782

PROJECT

DIFFUSION VARIATIONAL AUTOENCODER FOR TACKLING STOCHASTICITY IN MULTI-STEP REGRESSION STOCK PRICE PREDICTION

Prratham Kheskwani – 21D070051
Sahil Dharod – 210070026
Chanakya Varude – 210070092

TOPIC

Multi-step stock price prediction over a long-term horizon is crucial for forecasting its volatility, allowing financial institutions to price and hedge derivatives, and banks to quantify the risk in their trading books.

To tackle these issues, we combine a deep hierarchical variational autoencoder (VAE) and diffusion probabilistic techniques to do seq2seq stock prediction through a stochastic generative process

GOALS AND OBJECTIVES

Objective 1

We investigate the problem of generalization in the stock prediction task under the multi-step regression setting, and deal with stochasticity in both the input and target sequences.

Objective 2

We propose a solution that integrates a hierarchical VAE model, a stochastic diffusion process and a denoising component and implement it in an end-to-end model for stock prediction.

Objective 3

Using one-step denoising process can also be seen as removing the estimated aleatoric uncertainty resulting from data stochasticity.

01
FEATURE EXTRACTION

02
VAE STRUCTURE

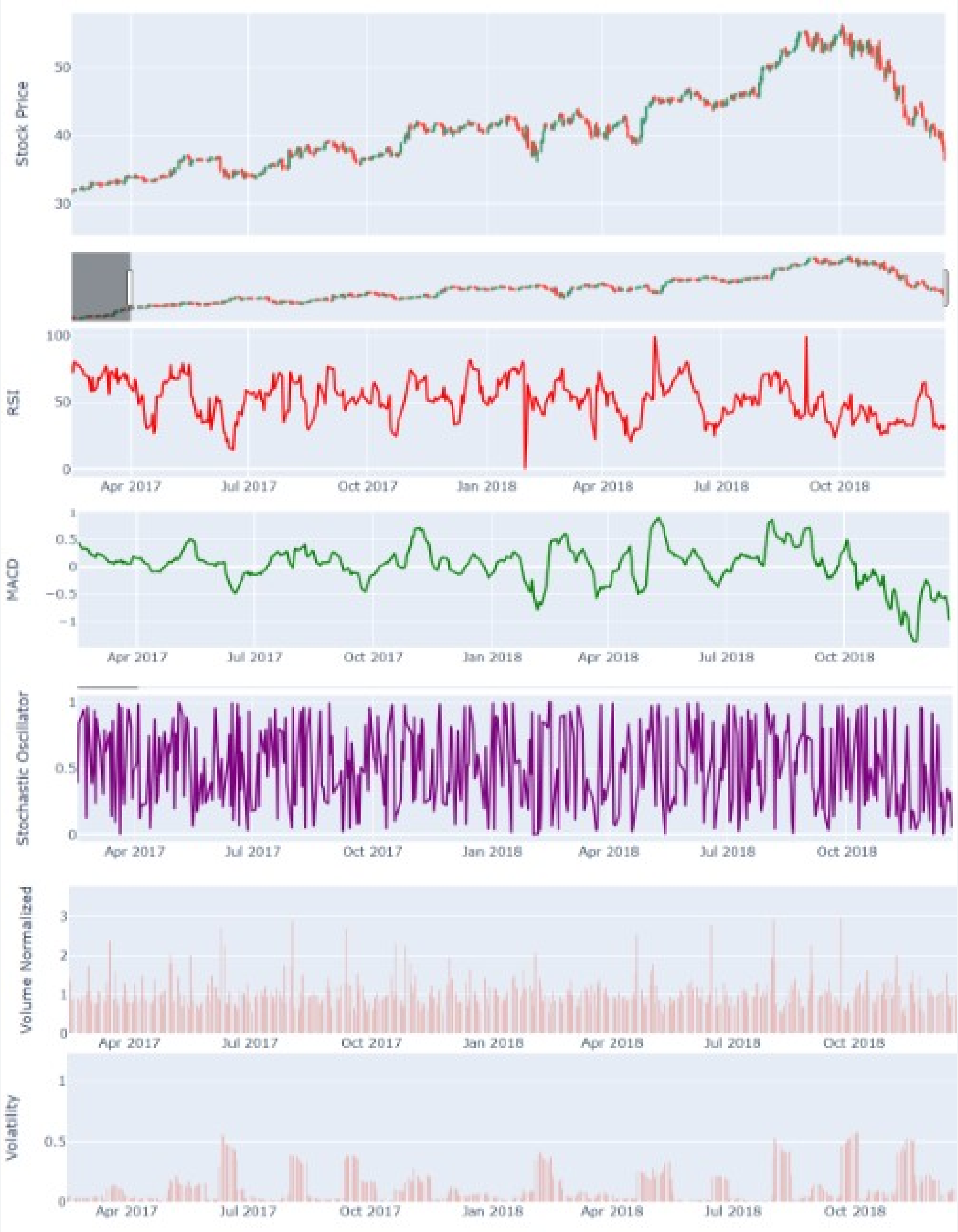
03
DIFFUSION MODELLING

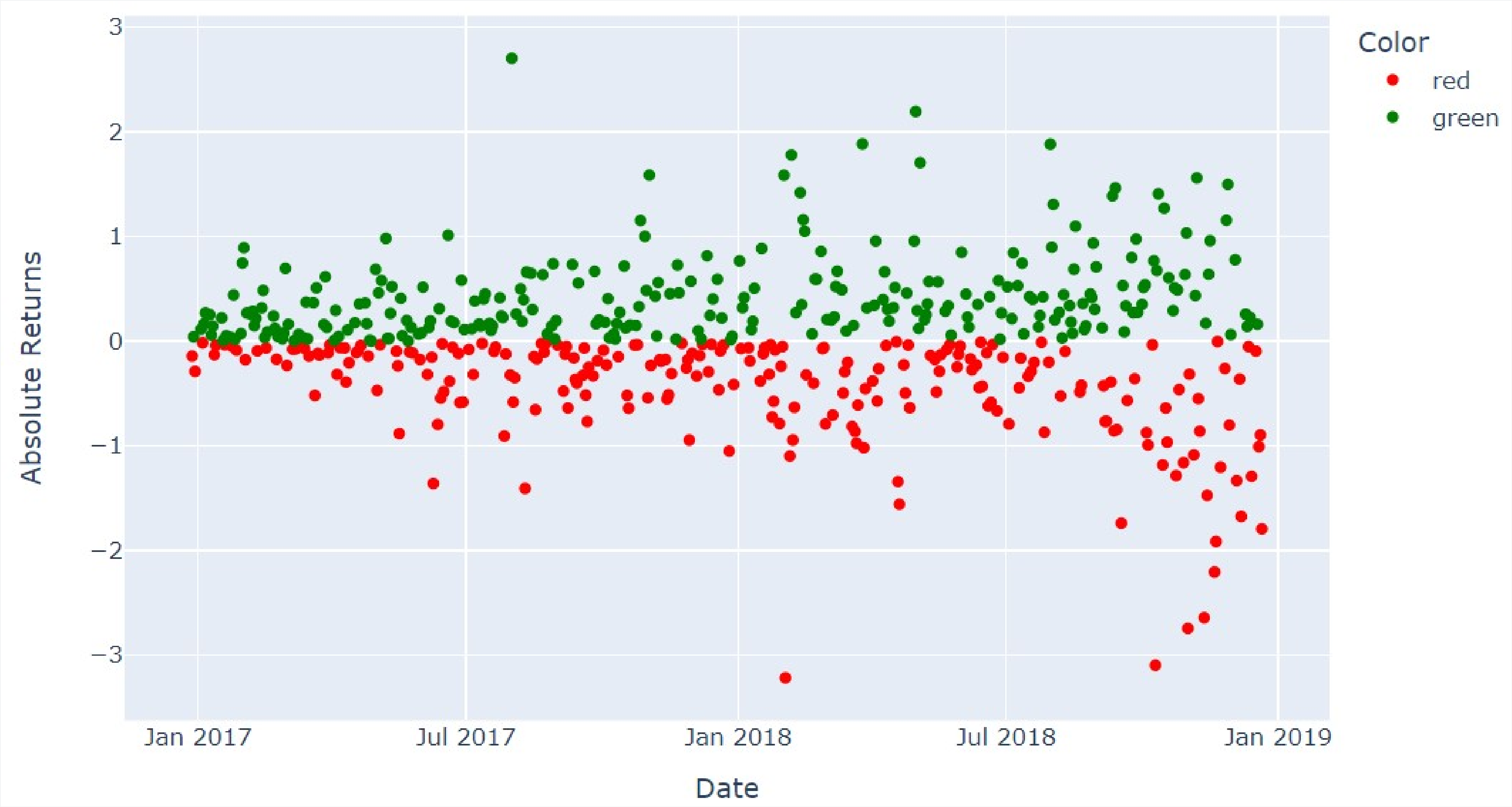
04
DENOISING STRUCTURE

FEATURE EXTRACTION

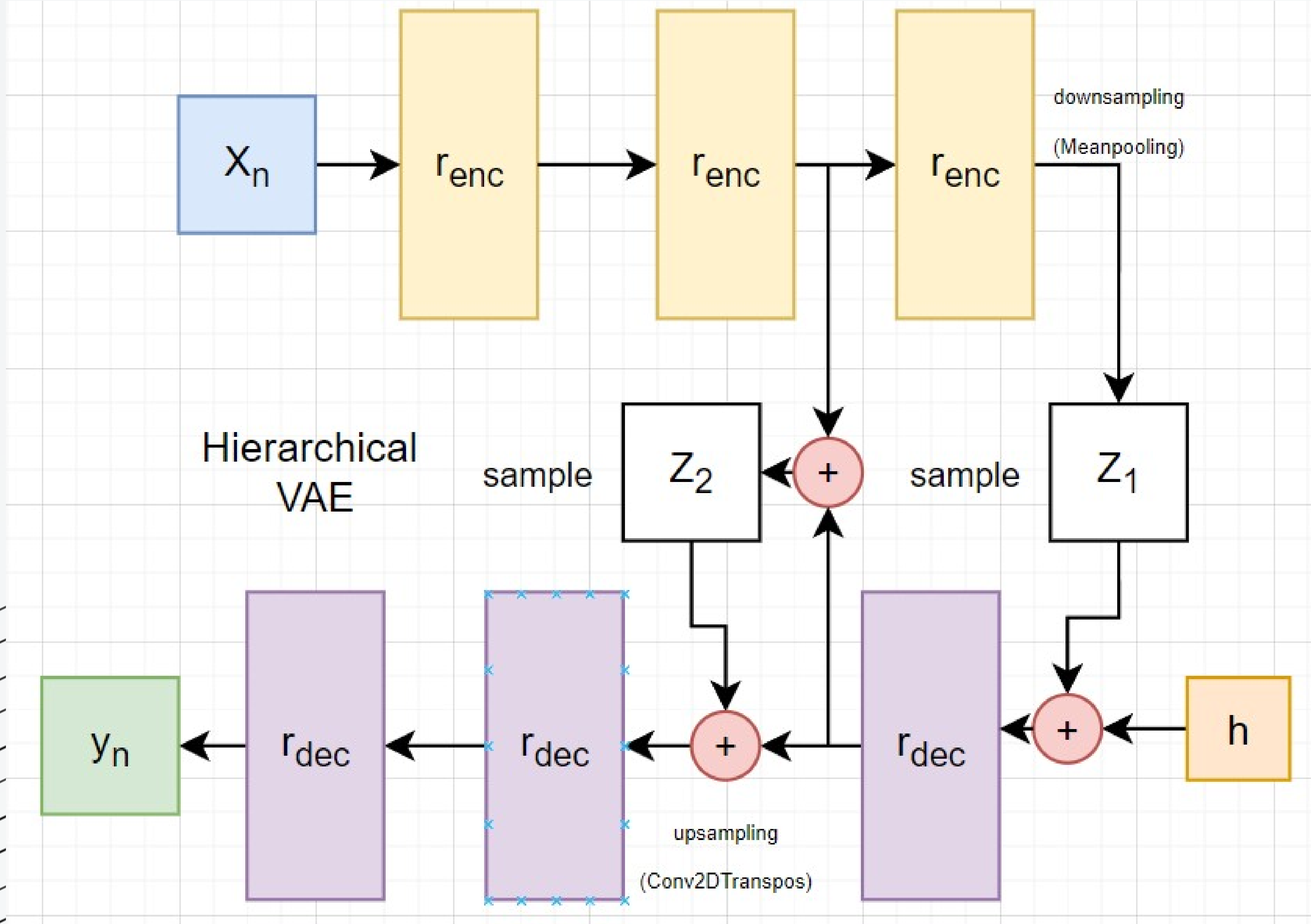
Features serve as the foundation on which models build their understanding of the data. High-quality and relevant features not only enhance model performance but also facilitate better generalization to unseen data.

Using derived features like Relative Strength Index, Volatility, Moving Average Convergence Divergence, Stochastic Oscillator, Periodic Wave terms apart from standard OHLCV


Stock Price
RSI
MACD
Stochastic Oscillator
Volume Normalized
Volatility
Apr 2017
Jul 2017
Oct 2017
Jan 2018
Apr 2018
Jul 2018
Oct 2018


Absolute Returns
Date
Jan 2017
Jul 2017
Jan 2018
Jul 2018
Jan 2019
Color
red
green

VAE STRUCTURE



Our primary model, the Noveau Variational AutoEncoder (NVAE) , initially designed for image generation, is repurposed as a seq2seq prediction model. The NVAE architecture involves generative and encoder networks with decoder and encoder residual cells. These cells employ batch normalization, Swish activation, convolution layers, and a Squeeze-and-Excitation (SE) layer, experimentally enhancing the VAE's performance. The decoder cells include depthwise separable convolution layers to capture long-range dependencies in the data while managing computational complexity.

DIFFUSION MODELLING

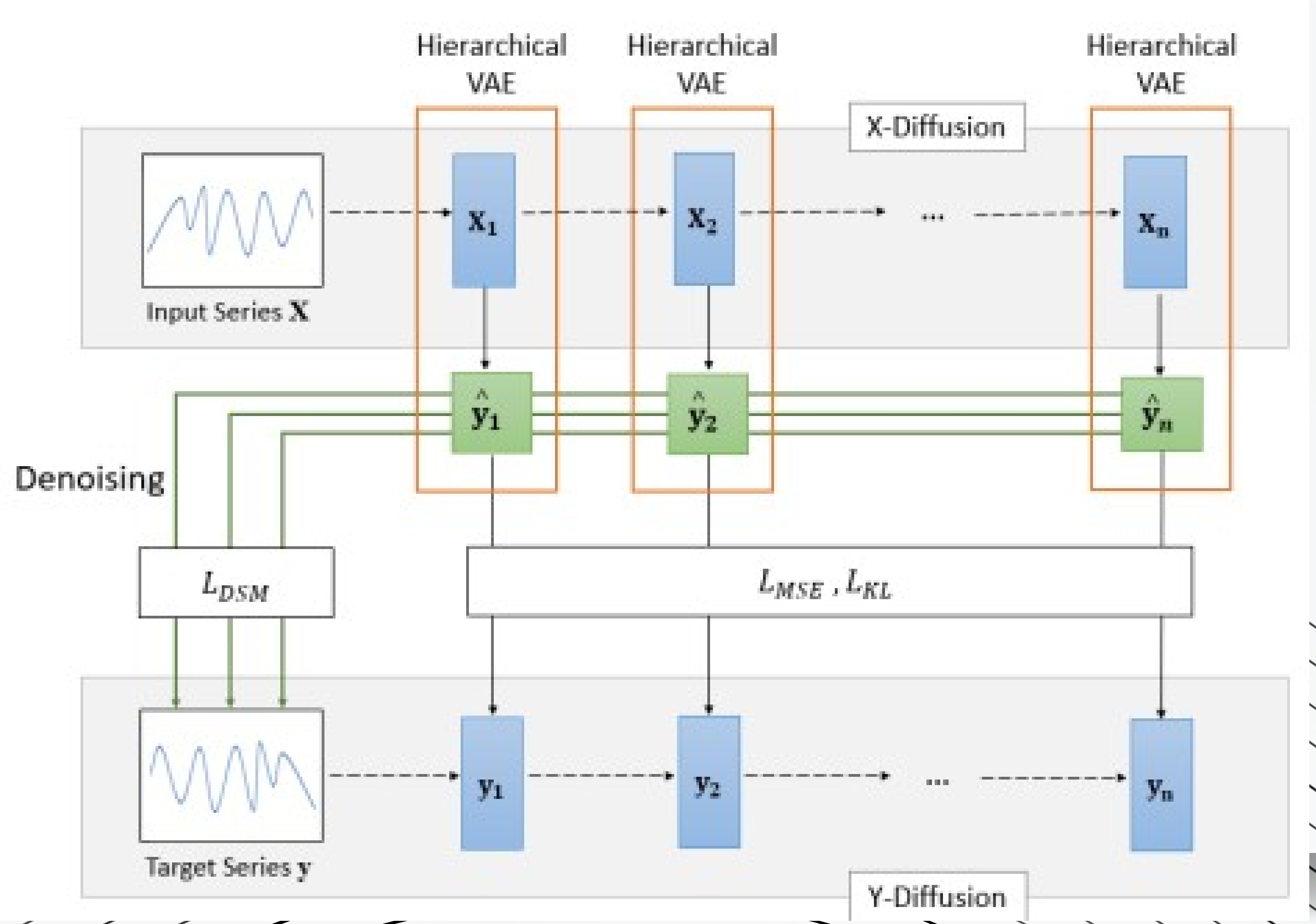
$$q(\mathbf{X}_{1:N}|\mathbf{X}) = \prod_{n=1}^{N} q(\mathbf{X}_n|\mathbf{X}_{n-1}),$$
$$q(\mathbf{X}_n|\mathbf{X}_{n-1}) = \mathcal{N}(\mathbf{X}_n; \sqrt{1-\beta_n}\mathbf{X}_{n-1}, \beta_n\mathbf{I});$$

$$q(\mathbf{X}_n|\mathbf{X}) = \mathcal{N}(\mathbf{X}_n; \sqrt{\bar{\alpha}_n}\mathbf{X}, (1-\bar{\alpha}_n)\mathbf{I}),$$
$$\mathbf{X}_n = \sqrt{\bar{\alpha}_n}\mathbf{X} + \sqrt{(1-\bar{\alpha}_n)}\boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(0, \mathbf{I});$$

$$q(\mathbf{y}_{1:N}|\mathbf{y}) = \prod_{n=1}^{N} q(\mathbf{y}_n|\mathbf{y}_{n-1}),$$
$$q(\mathbf{y}_n|\mathbf{y}_{n-1}) = \mathcal{N}(\mathbf{y}_n; \sqrt{1-\beta'_n}\mathbf{y}_{n-1}, \beta'_n\mathbf{I}).$$
$$q(\mathbf{y}_n|\mathbf{y}) = \mathcal{N}(\mathbf{y}_n; \sqrt{\bar{\alpha}'_n}\mathbf{y}, (1-\bar{\alpha}'_n)\mathbf{I}),$$
$$\mathbf{y}_n = \sqrt{\bar{\alpha}_n}\mathbf{y} + \sqrt{(1-\bar{\alpha}_n)}\boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(0, \mathbf{I}).$$

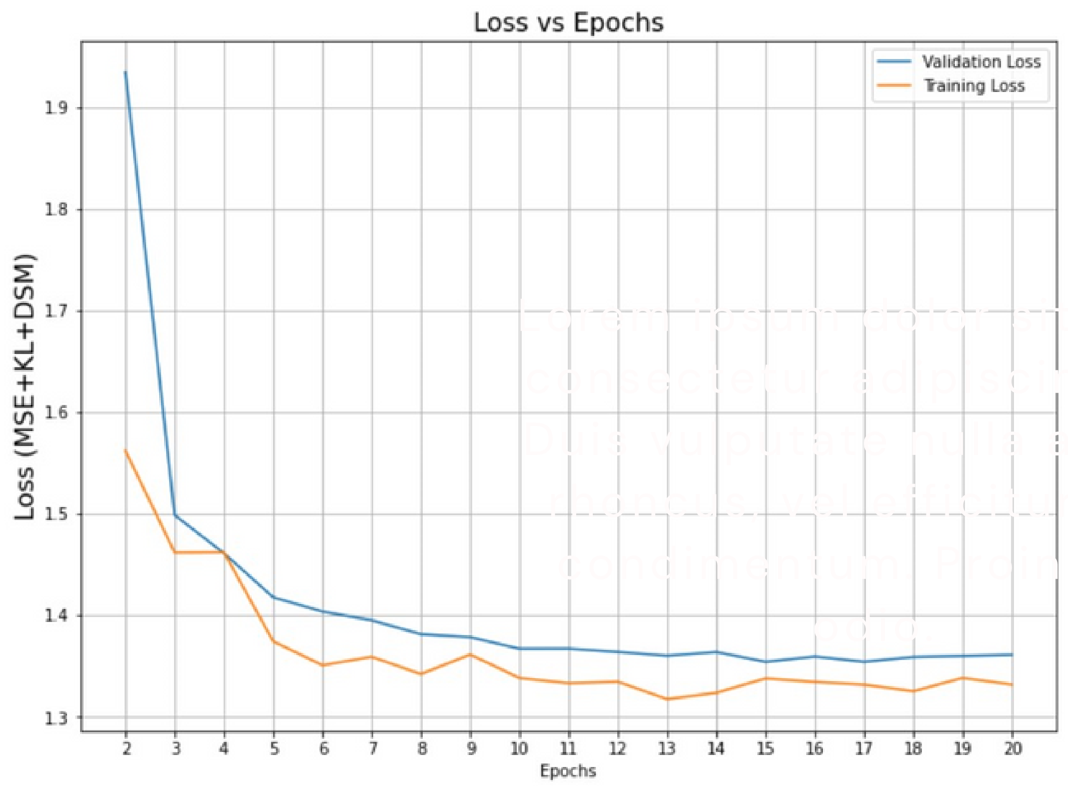
1) By adding diffusion noise to both X and y sequences and aligning distributions from a generative model and the diffusion process, overall uncertainty in the generative model and inherent data noise (aleatoric uncertainty) reduces
2) The coupling of the generative and diffusion processes minimizes prediction uncertainty, reducing divergence between noisy and original sequences.
3) Introducing coupled Gaussian noise to target sequences generates noisy samples aiding in simulating target series' stochasticity, reducing uncertainty in the predicted sequence.

DENOISING STRUCTURE

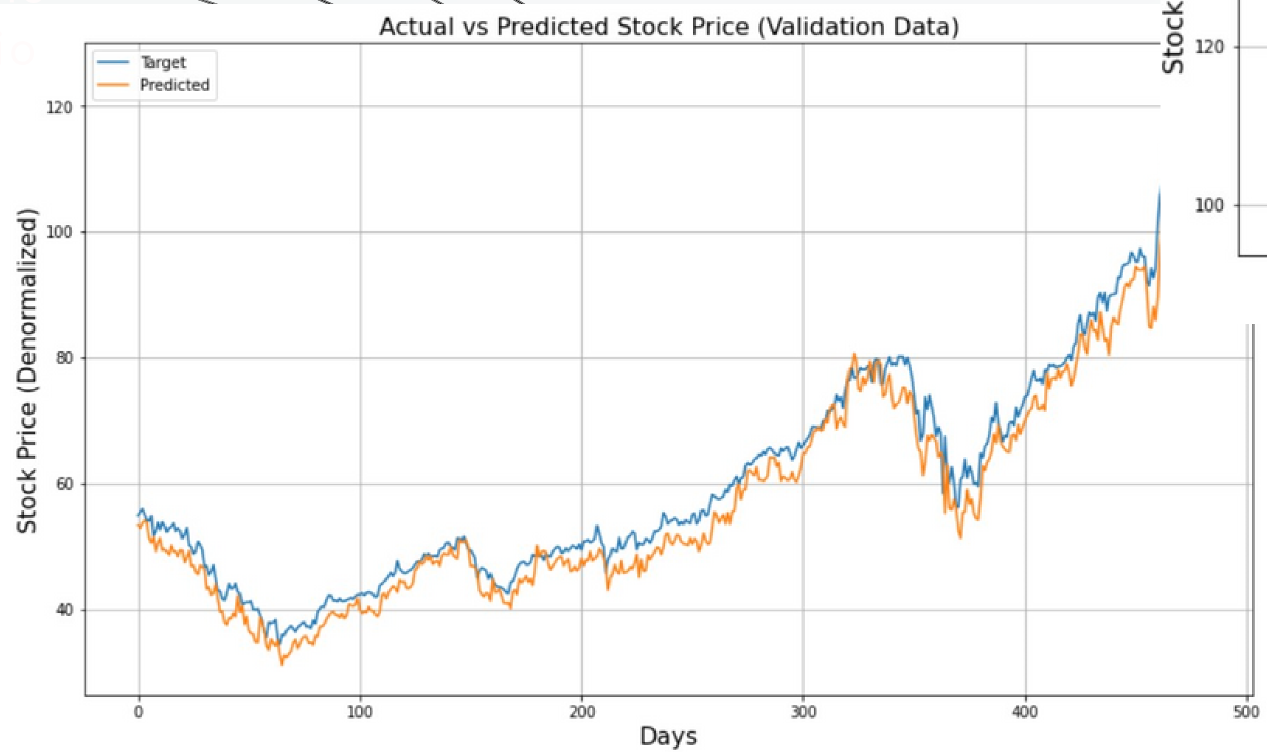


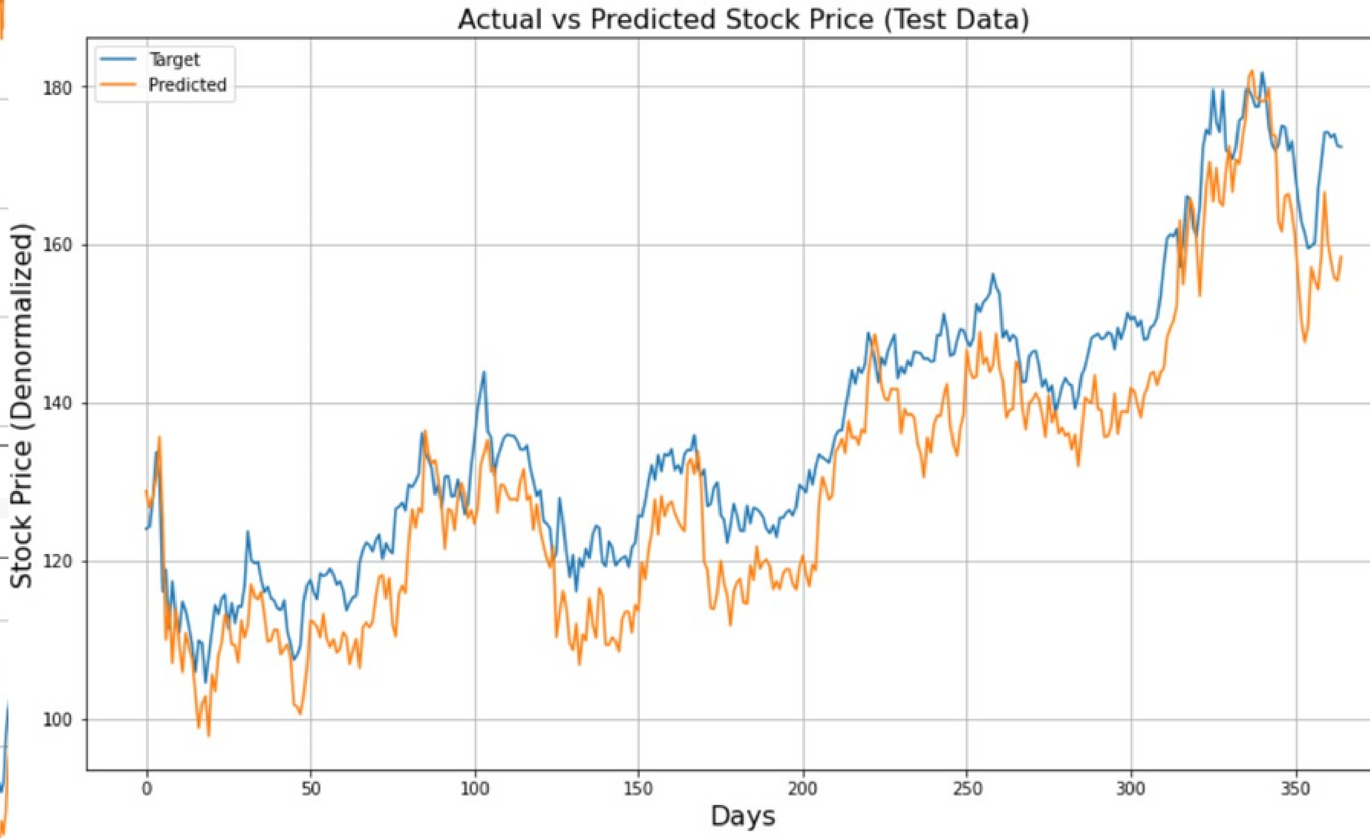
Denoising score-matching (DSM) process, adapted from standard diffusion probabilistic models, matches gradients to learn an energy function, aiding in recovering yr from a corrupted y sequence with Gaussian noise. At test-time, a one-step denoising jump is executed to obtain the final predicted sequence ŷfinal, effectively reducing estimated aleatoric uncertainty resulting from data stochasticity.

RESULTS


Loss vs Epochs
Validation Loss
Training Loss
Loss (MSE+KL+DSM)
Epochs


Actual vs Predicted Stock Price (Training Data)
Target
Predicted
Stock Price (Denormalized)
Days


Actual vs Predicted Stock Price (Validation Data)
Target
Predicted
Stock Price (Denormalized)
Days


Actual vs Predicted Stock Price (Test Data)
Target
Predicted
Stock Price (Denormalized)
Days

CHALLENGES FACED

UNDERSTANDING

1) Understanding the architecture, and implementing it in our own way

DEBUGGING

2) Dimensionality matching while building different constituent models

3) Dimensionality matching faced in the training Loop

4) Other errors given while construction of the model

OPTIMAL BEHAVIOUR

4) Unpredictable behaviour of losses, which we fixed using Xavier initialization and the use of learning rate schedulers.

5)Training Multiple times to attain the best possible results

THANK'S FOR WATCHING