

Configuration Manual

MSc Research Practicum
MSc Data Analytics

Vasudha Kurugundra
Student ID: x23305631

School of Computing
National College of Ireland

Supervisor: Shubham Subhnil

National College of Ireland
MSc Project Submission Sheet
School of Computing



Student Name: Vasudha Kurugundra
Student ID: X23305631
Programme: MSc Data Analytics **Year:** 2024-2025
Module: MSc (Research)Practicum
Lecturer: Shubham Subhnil
Submission Due Date: 11/08/2025
Project Title: Enhancing Financial Synthetic Data Generation Through Local Historical Pattern Retrieval

Word Count: 246 **Page Count:** 4

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: Vasudha Kurugundra
Date: 11/08/2025

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	<input type="checkbox"/>
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	<input type="checkbox"/>
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	<input type="checkbox"/>

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	

Enhancing Financial Synthetic Data Generation Through Local Historical Pattern Retrieval

Vasudha Kurugundra

Student Id:x23305631

1 Introduction

This configuration manual provides a comprehensive guide to setting up, training, and running inference for the **S&P 500 Stock Market Prediction Project**, which leverages **Generative Adversarial Networks (GAN)** and **Variational Autoencoders (VAE)** with Retrieval-Augmented Generation (RAG) enhancements.

The document is designed to help users replicate the experimental setup used in this research. It includes detailed environment setup instructions, software dependencies, dataset preparation, model training commands, and inference procedures.

2 Hardware & Software Requirements

Minimum Requirements:

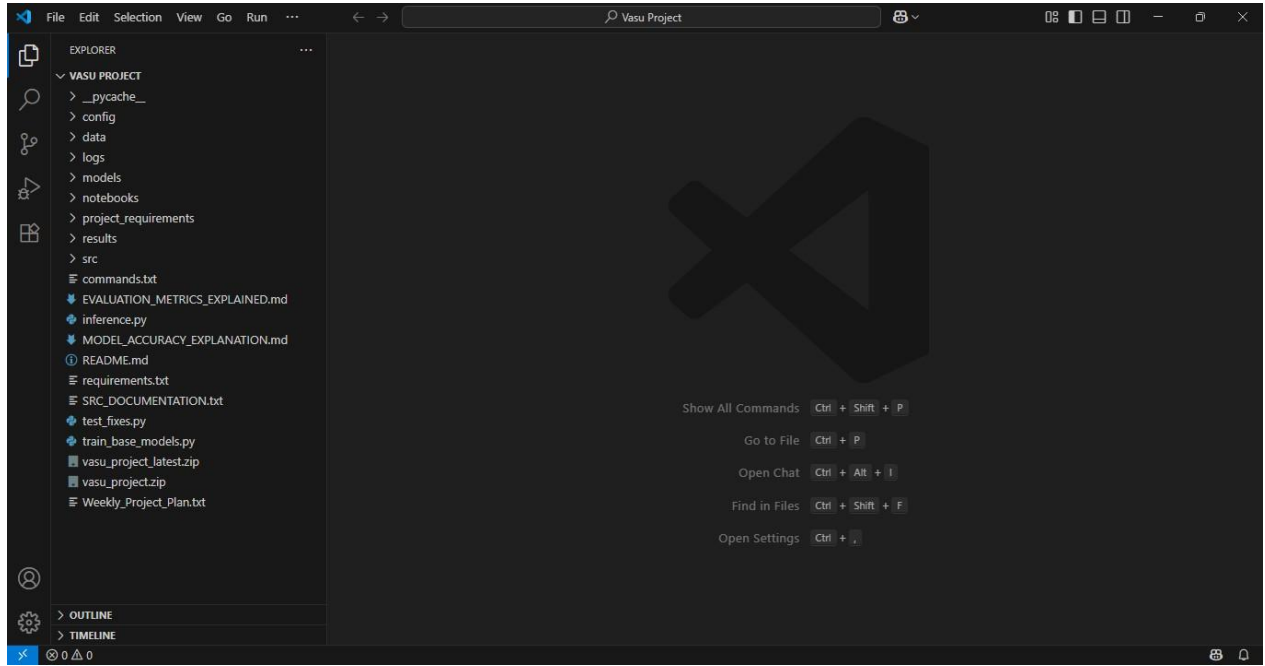
- **Processor:** Quad-core CPU
- **RAM:** 16 GB
- **Storage:** 50 GB free space
- **GPU:** NVIDIA GPU with 8 GB VRAM (e.g., GTX 1080 or higher)

Recommended Setup:

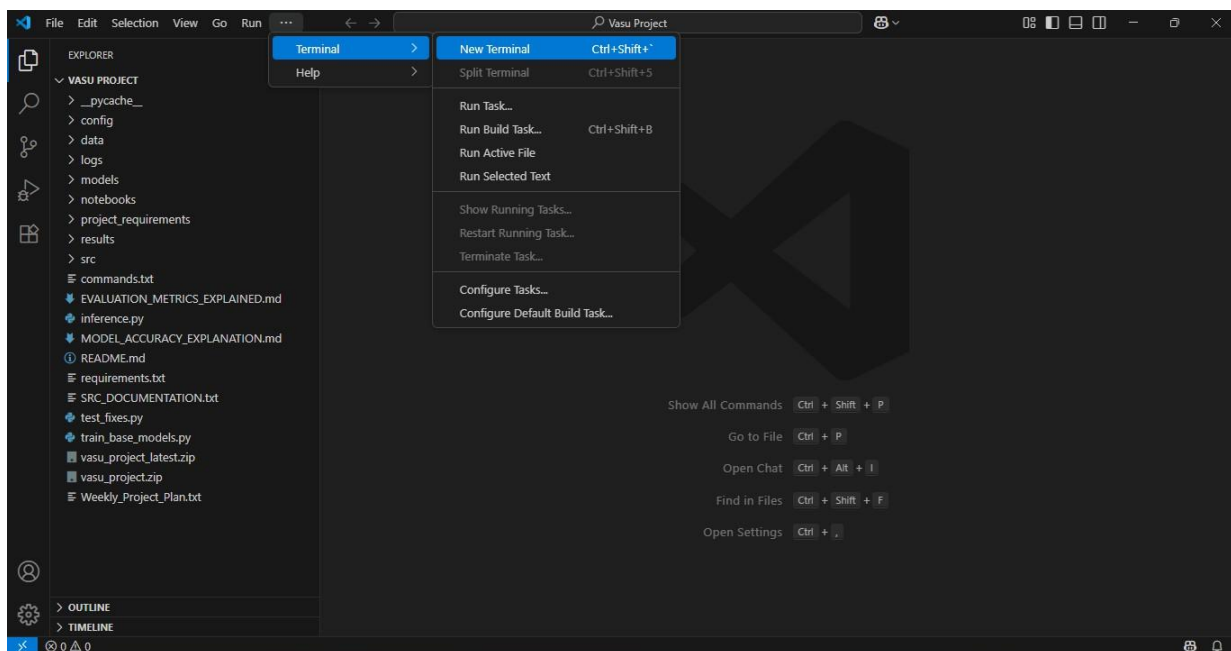
- **Processor:** 8-core CPU
- **RAM:** 32 GB
- **Storage:** 200 GB SSD
- **GPU:** NVIDIA RTX 3090 / A100 for accelerated training
- **Operating System:** Windows 10 / Ubuntu 20.04 LTS
- **Python Version:** 3.10 (via Conda)

3 Environment Setup

3.1 Download & Unzip the folder to open in VSCode



3.2 Click on the new terminal in the top right hand side



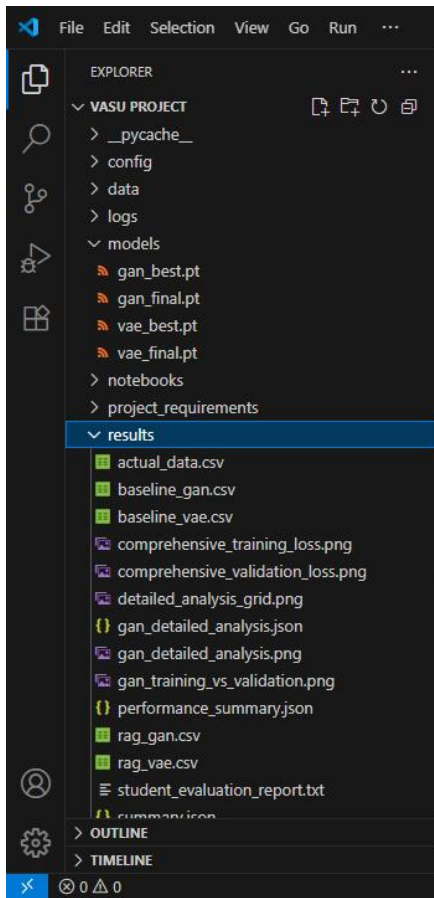
3.4 Run the following commands in an order

- conda create --name my_project python=3.10 -y
- conda activate my_project
- python train_base_models.py --model both --epochs 100 --start-date 2010-01-01

```
33 from src.evaluation.plot_utils import plot_loss_and_accuracy
34
35 # Add src to path
36 sys.path.append(os.path.join(os.path.dirname(__file__), 'src'))
37
38 # Import our modules
39 from src.data_utils import FinancialDataProcessor, load_and_process_data
40 from src.models.base_gan import BaseFinancialWGAN
41 from src.models.base_vae import BaseFinancialVAE
42 from src.evaluation.metrics import EnhancedModelEvaluator
43 from src.fetch_sp500 import fetch_sp500_data, save_data_to_csv
44 from src.hyperparameter_tuning import EarlyStopping, optimize_hyperparameters
45
46 # Set up logging
47 logging.basicConfig(level=logging.INFO, format='%(asctime)s - %(levelname)s - %(message)s')
48 logger = logging.getLogger(__name__)
49
50
```

```
PS C:\Users\kurug\Desktop\VASU PROJECT> python train_base_models.py --model both --epochs 100
INFO:__main__:Initialized Enhanced BaseModelTrainer on cpu
INFO:__main__:Training epochs updated: 100
INFO:__main__:Preparing data for training
INFO:__main__:Fetching S&P 500 data from 2010-01-01 to 2025-12-31
Fetching S&P 500 data from 2010-01-01 to 2025-12-31
Successfully fetched 3924 records
Error saving data to CSV: [WinError 183] Cannot create a file when that file already exists: 'data/sp500_data.csv'
INFO:__main__:Data fetched and saved: 3924 records
INFO:src.data_utils:Initialized FinancialDataProcessor with 5 features
INFO:src.data_utils:Sequence length: 30, Normalize: True
```

3.5 Once the model training is done, you can see the following folders created



3.6 Run the following commands to inference using the trained model

a. `python inference.py --query "Generate data for volatile market conditions" --samples 1500`

b. Checkout the below files getting created in the results folder

