Getting Started with the SWAN-SF Data Analysis

Welcome to our GitHub repository! This guide will help you set up and run the code effectively. Our project involves multiple data preprocessing steps, so it's essential to follow the sequence correctly.

Prerequisites

Before you start, make sure you have the following:

- **SWAN-SF Dataset**: Download it from Harvard Dataverse (https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/EBCFKM).
- **Python Packages**: Ensure you have these packages installed: `pandas`, `numpy`, `matplotlib`, `seaborn`, `tensorflow`, `tqdm`, `pickle`, `sklearn`, `scipy`, `imblearn`. The code for `timegan` is included in the repository, so no additional installation is required for this package.

Setting Up Your Environment

1. **Directory Setup**: Modify the following lines in the source code to match your system's directory structure:

```
data_dir = "<Your path>/SWANSF/Downloaded_Data/"
data_dir_save = "<Your path>/SWANSF/code/"
```

2. **Sequential Execution**: Start from Notebook 1 and proceed in order. Each notebook relies on the data prepared in the previous steps.

Notebooks Overview

- **Notebook 1**: Reads SWAN-SF samples and combines them into a single `.pkl` file (time series samples) and a `.csv` file (labels for each partition).
- **Notebook 2**: Focuses on Missing Value Imputation, utilizing data from **Notebook 1**.
- **Notebook 3**: Centers on Near Decision Boundary Sample Removal.
- **Notebook 4 & 5**: Concentrate on Normalization.
- **Notebook 6**: Offers Visualizations of the dataset.
- **Notebook 7 & 8**: Implement Classification using eight classifiers.
- **Notebook 9**: Applies Over-sampling techniques.
- **Notebook 10**: Combines Over- and Under-sampling techniques.
- **Notebook 11, 12, & 13**: Apply preprocessing techniques post-sampling (Normalization).

- **Notebook 14, 15, 16, & 17**: Implement Classification using eight classifiers after Sampling.
- **Notebook 18**: Presents Final Visualizations.

Copy Right
