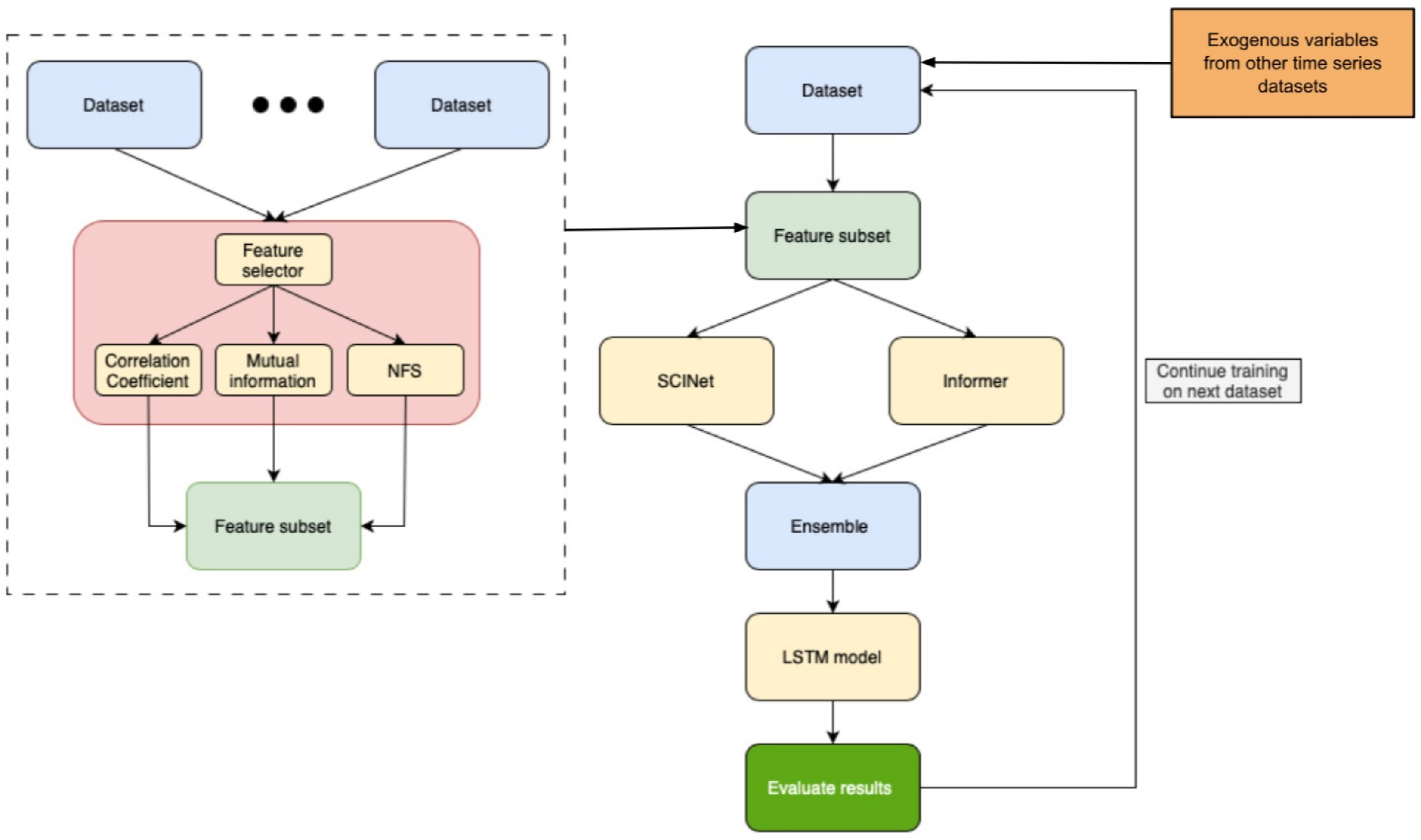
NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE
School of Computer Science and Engineering
College of Engineering

Multivariate Earthquake Prediction

Student: Vaidyanathan Abhishek Supervisor: Jagath C Rajapakse

Motivation and Proposed Approach:

Earthquake prediction has been an important field of study, especially with respect to disaster mitigation strategies. Previous approaches study the use of different techniques such as machine learning and deep learning techniques to model complex time series data. This research proposes a novel architecture that utilises the complex temporal and spatial dependencies along with exogenous signals for adequate prediction of earthquakes.



Data Statistics:

This research makes use of time series data available on AETA earthquake dataset. We make use of data collected from around 100 sensors placed across the south China sea. The data comprises of electromagnetic and geo-acoustic features.

Pipeline:

The proposed pipeline makes use of feature selection methods such as rule based and learning based algorithms. Furthermore, state of the art model utilising attention-based mechanisms were experimented to model the complex dependencies within time series data.

Experimental Results:

Our research involves the use of state-of-the-art model such as SCINet and Informer models. The two models consist of attention-based mechanisms that capture temporal or spatial dependencies within time series data. We experiment with neural network pruning algorithms for feature selection as well as standard rule-based algorithms such as correlation coefficient and mutual information to get the optimal feature subset. Final experiments were conducted using a window of 96 time steps and obtained predictions for 48 time steps forward. Results for different experiments are as shown below.

Model [Datasets]	Correlation FS	Mutual Information FS	NFS
SCINet [10]	0.09	0.15	0.12
Informer [10]	0.10	0.08	0.14
LSTM Ensemble [10]	0.18	0.17	0.13
Random Forest Ensemble [10]	0.06	0.03	0.10
SCINet [20]	0.12	0.10	0.13
Informer [20]	0.10	0.09	0.13
LSTM Ensemble [20]	0.16	0.19	0.17
Random Forest Ensemble [10]	0.06	0.03	0.12

Conclusion:

The table shows that optimal results are obtained from the random forest ensemble. However, on visualizing, the LSTM model produces results with the best generalization ability while Random forest predicts most values around the center value. This leads to infer that relying solely on losses may not always suffice to describe model efficacy.