From FogNets to Neural Nets

Texas Tech University |

Final Project Report

Vivekanand Praturi

[Year]

Contents

1.Introduction

1.1 Problem statement

2.Data Visualization

3.Data Mining and Cleaning

4.Approach and Algorithms

5.Application of Multivariate tools and Techniques

6.Results

7.Future Work

8.Literature

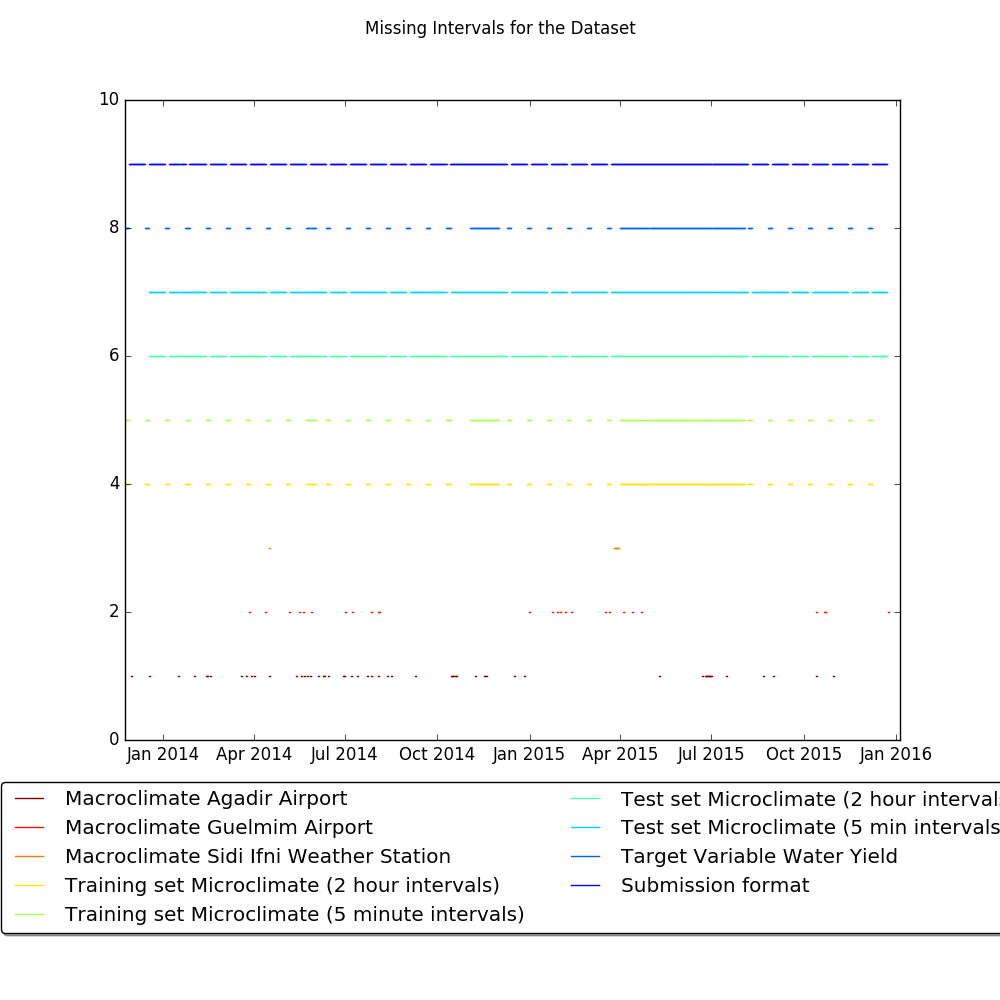
**1.Introduction**

* 1. **Problem Statement:**

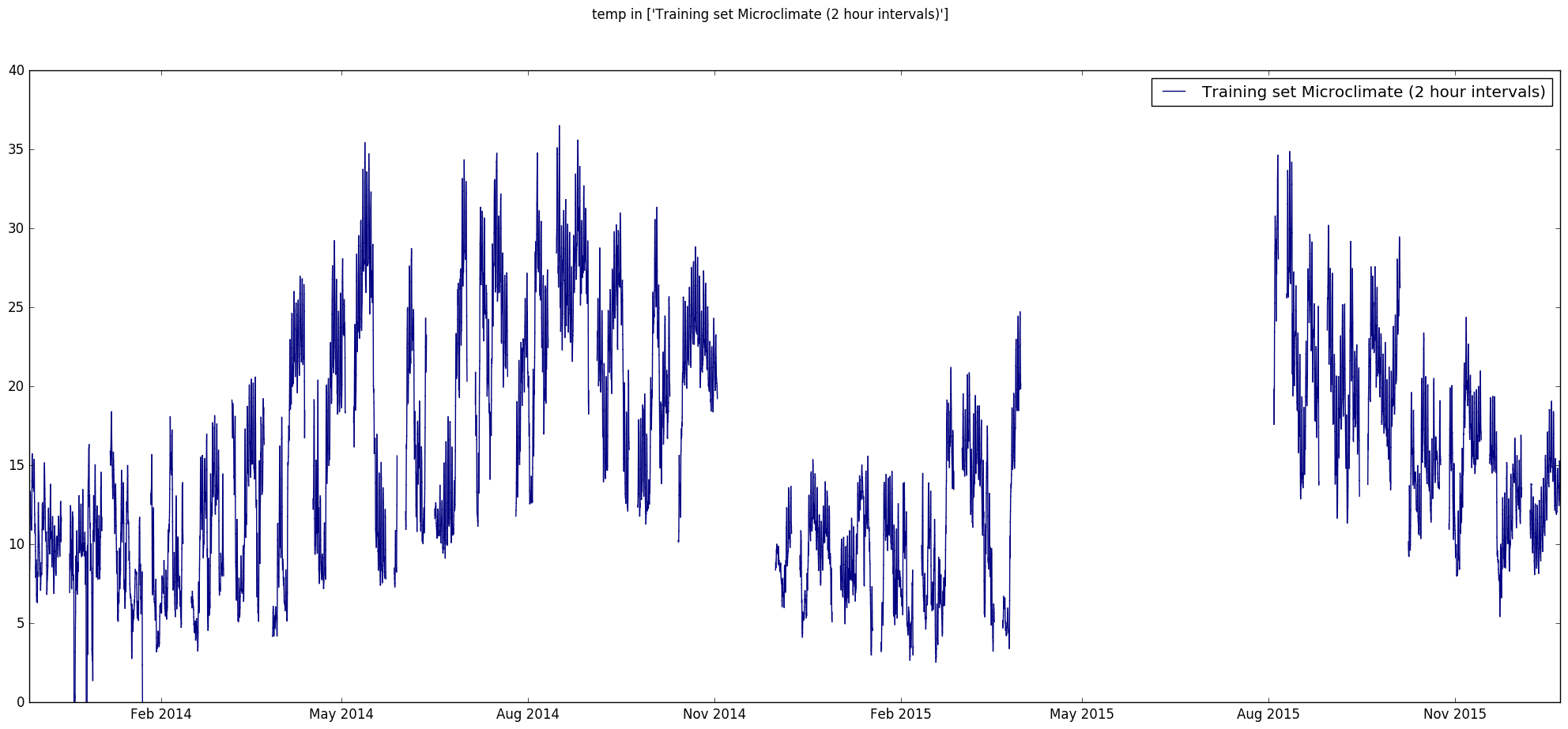
On April 15, 1912, during her maiden voyage, the Titanic sank after colliding with an iceberg, killing 1502 out of 2224 passengers and crew. The objective of this task is to complete the analysis of what sorts of people were likely to survive in the sinking of the RMS Titanic is one of the most infamous shipwrecks in history. The need to address this problem is important to avoid such massive tragedies and ensure safety regulations for ships.

**2.Data Visualization:**

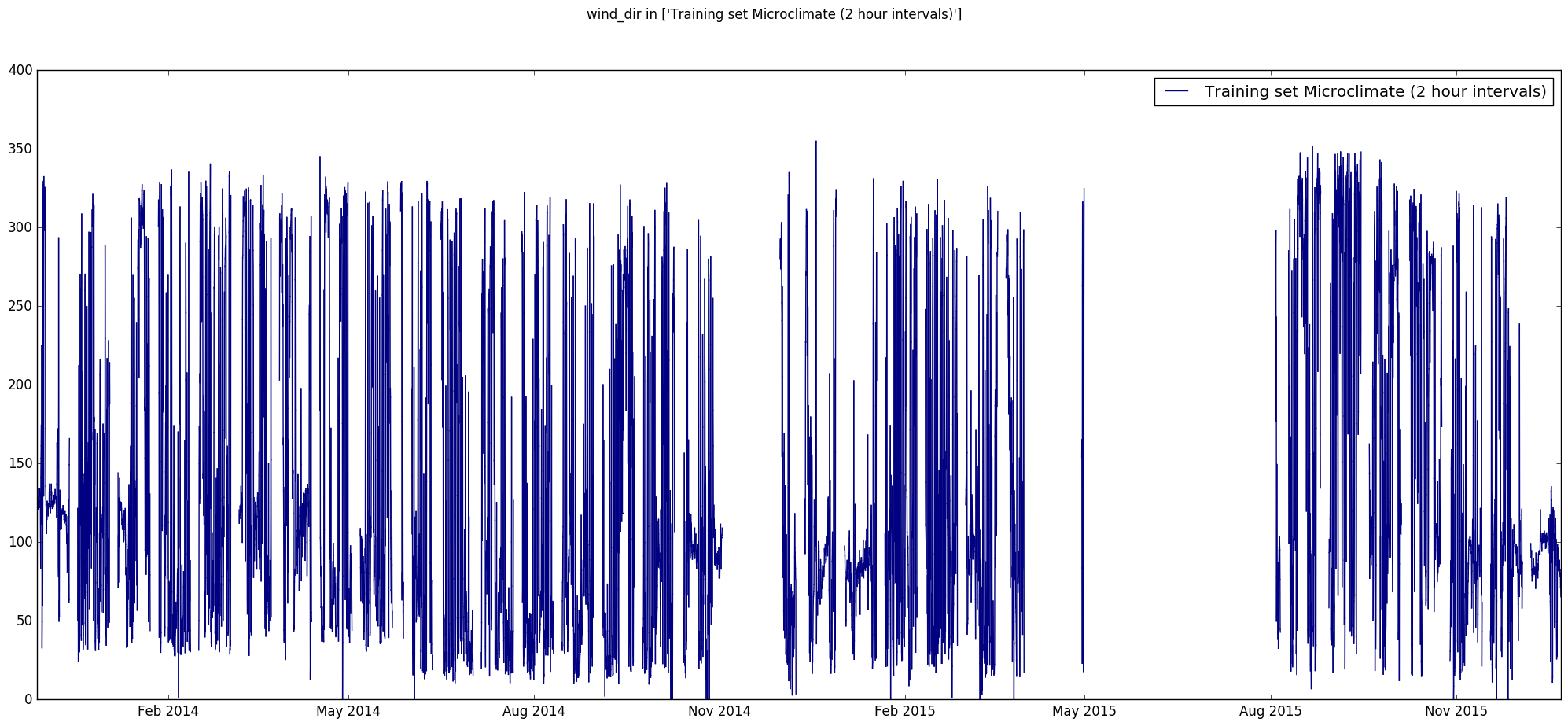
**Missing Values Plot:**

****

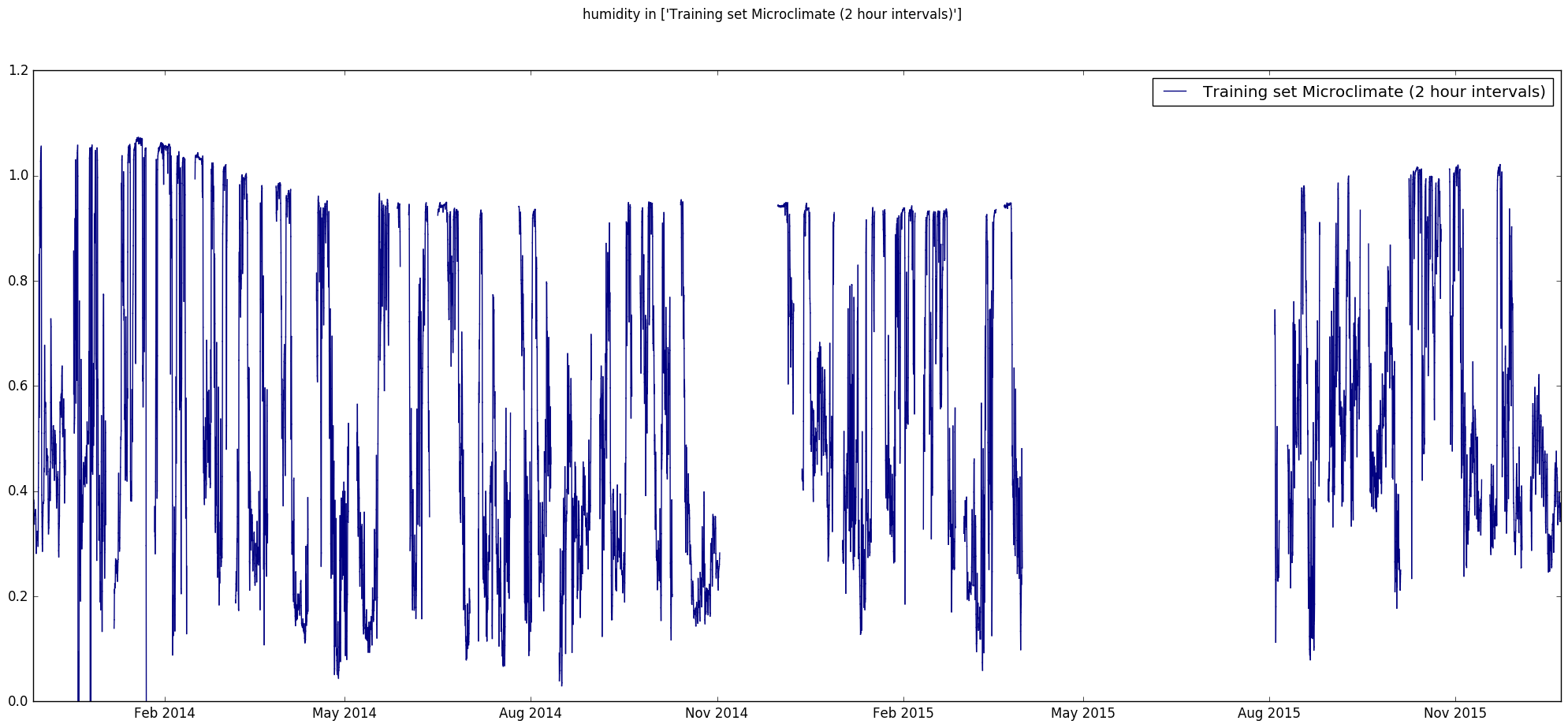
**Micro Features Plot**

**Temperature: **

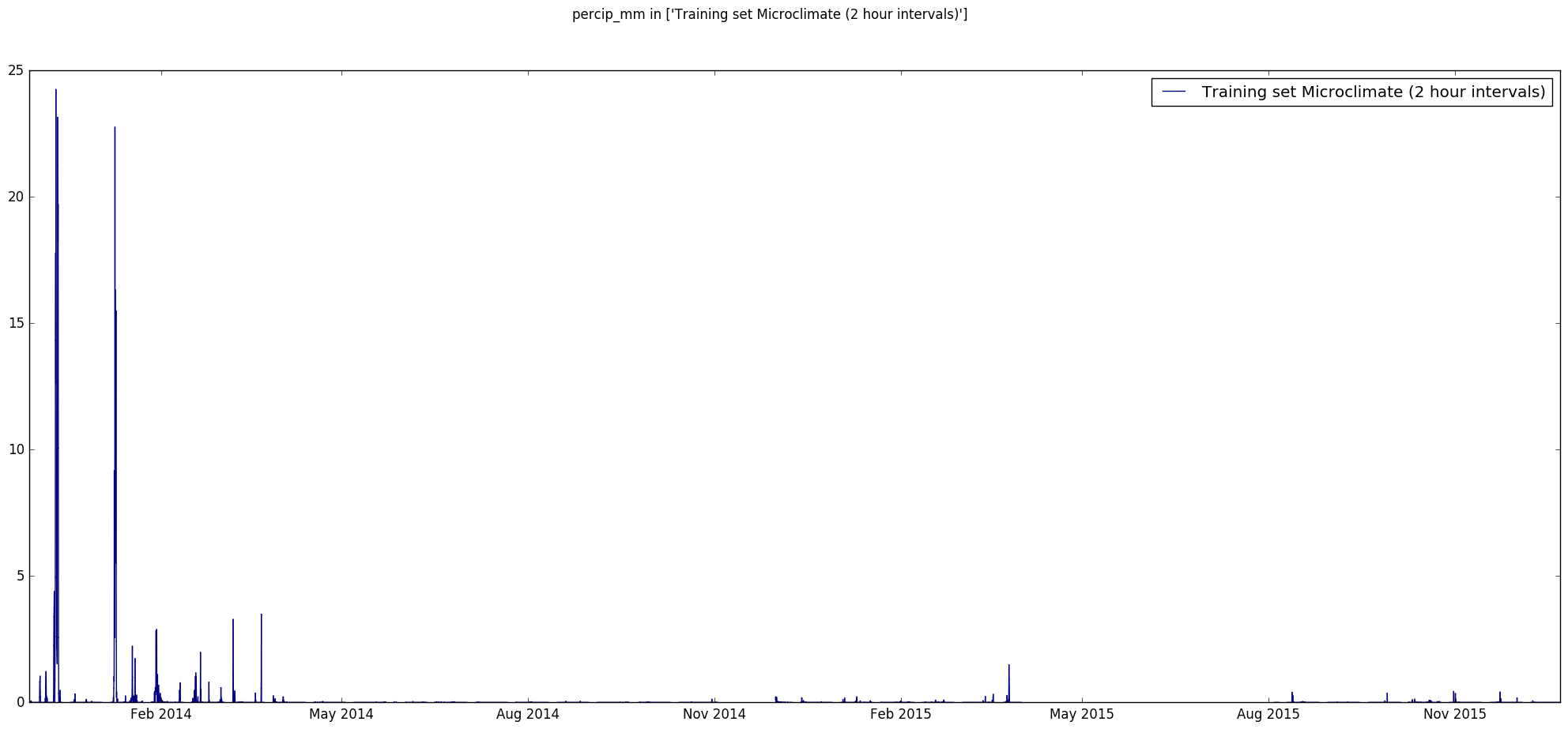
**Wind Direction:**

****

**Humidity:**

****

**Precipitation:**

****

**3.Data Mining and Cleaning**

* Added Day,Hour,Month,Year variables along with Lagged Variables and Moving averages to account for trend.
* Imputing Missing Data :
* Time Series Imputation using ARIMA and Kalman Filters (Package--TS Impute)
* Mice Imputation
* Integrating imputed 5min dataset to augment 2hr FogNet data in prediction

**4. Approaches and Algorithms**

* Converted Time series data to Dichotomous Variables and Added Lagged Variables and Moving averages to account for trend.
* Used Fognet Data with out Engineered variables and used the predictions in the ensembles along with predictions of Yield from Fognet + Nearby Airport Data
* Using different feature sets with different Algorithms
* Used higher weights for predictions Fognets in enesmbles

**5.Application of Multivariate tools and Techniques:**

* H2O : Deep Learning
* H2O : GBM
* H2O : GLM
* H2O : Random Forest
* Ensemble Methods

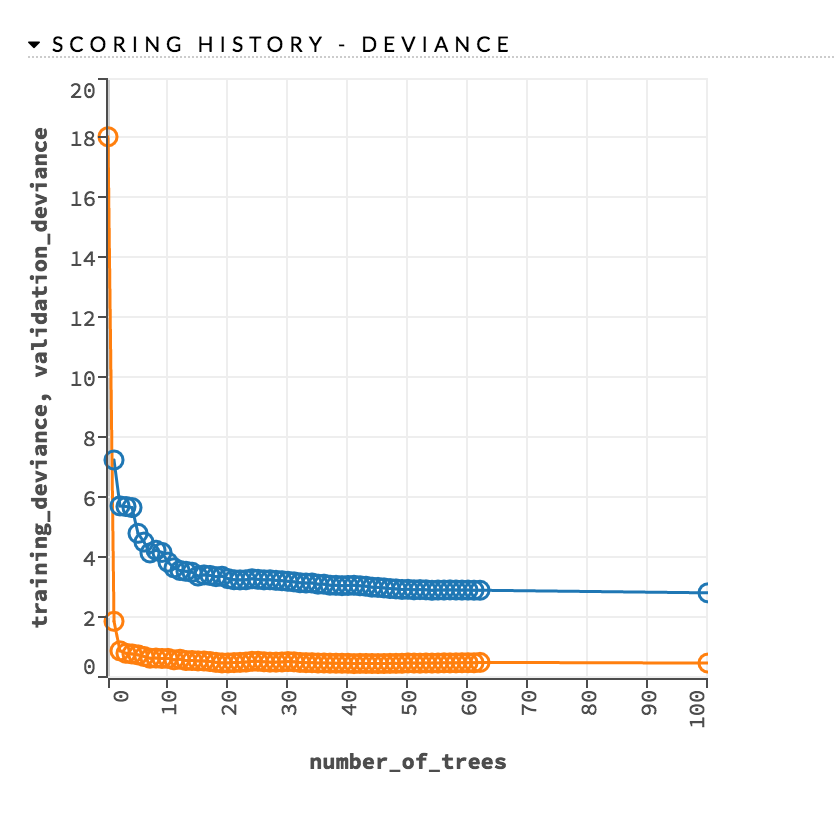
Used GBM, GLM, Random Forests as base learners with GBM meta learner.

The models are given random parameters based but there is a room for improvement by choosing the right parameters. So choosing the right parameters for the models is key to get better results.

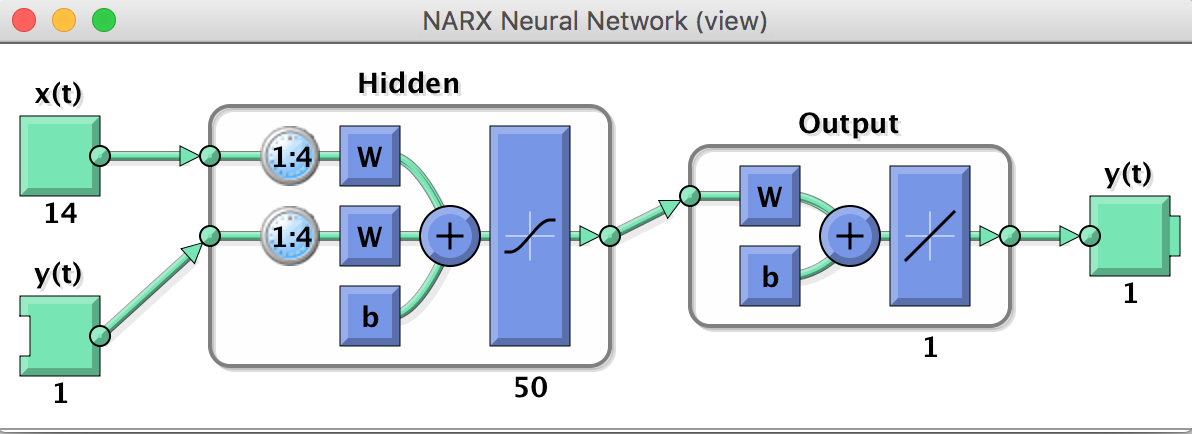
* Hyper parameter tuning with Grid Search

The model takes the range of parameters and generates all possible models and returns the result for all the models. The best model can be selected by sorting the models based on the Mean square error or Cross Validation error.

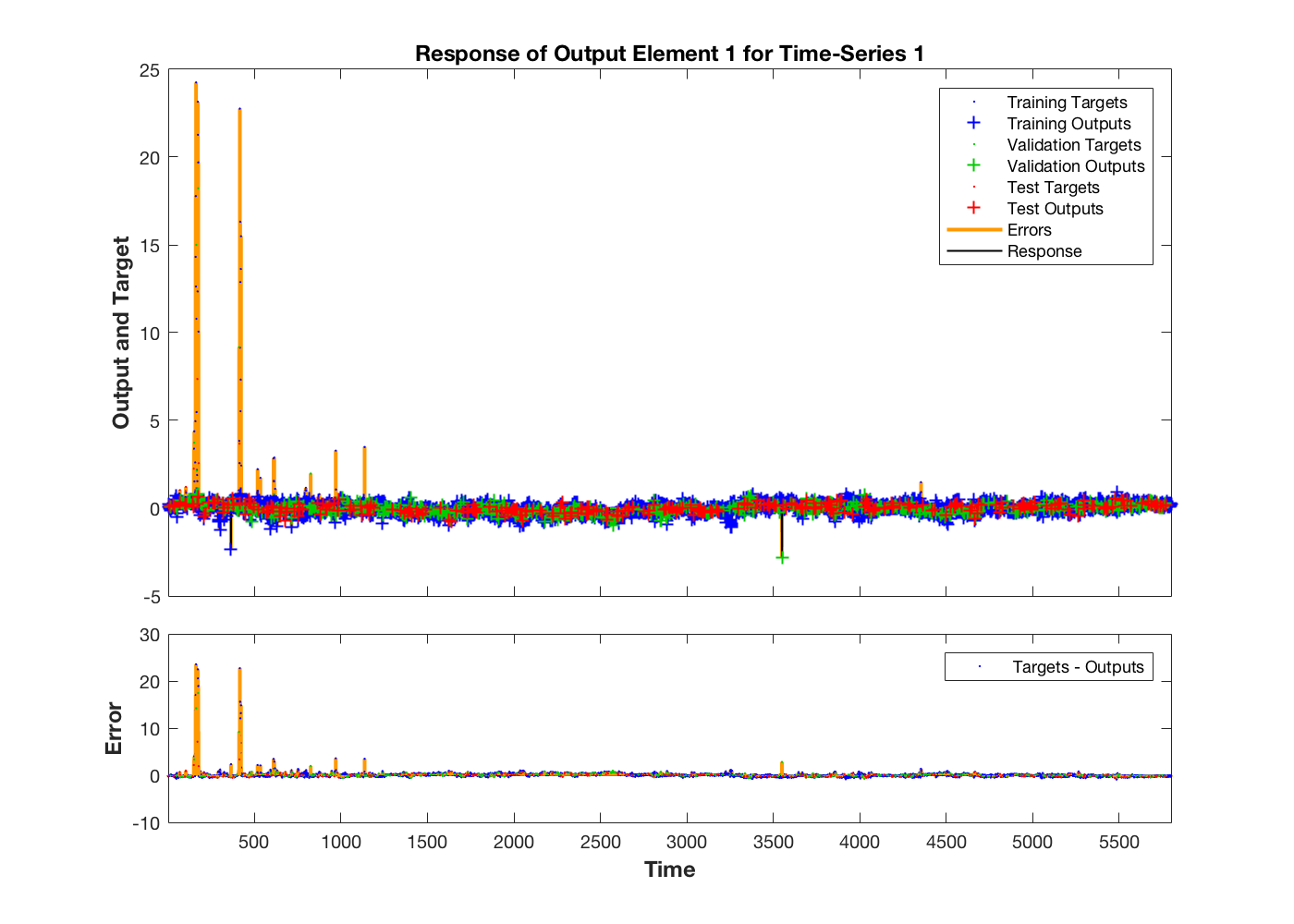
**Random Forest Model performance:**

****

**Narx Neural Networks:**

****

*y*(*t*)*f*(*y*(*t*−1)*y*(*t*−2)*y*(*t*−*ny*)*u*(*t*−1)*u*(*t*−2)*u*(*t*−*nu*))

****

**6.Driven Data Submission Result:**

****

**7.Future Work**

**Using following algorithms for time series analysis**

* **Tuning NARX Networks**
* **Recurrent Neural Networks**
* **Time Delay Networks**

**8.Literature**

**Why Lagged Dependent Variables Can suppress the explanatory power of other Independent Variables**

* [**https://www.princeton.edu/csdp/events/Achen121201/achen.pdf**](https://www.princeton.edu/csdp/events/Achen121201/achen.pdf)

**Using Neural Nets in Time Series Prediction**

* **http://www.maths.bath.ac.uk/~jjf23/papers/neural/nnts.pdf**

**Time series Prediction and Neural Networks**

* [**https://uhra.herts.ac.uk/bitstream/handle/2299/593/102081.pdf**](https://uhra.herts.ac.uk/bitstream/handle/2299/593/102081.pdf)

**Imputation of time series data**

* [**http://www.et.bs.ehu.es/~etptupaf/pub/papiros/mits.pdf**](http://www.et.bs.ehu.es/~etptupaf/pub/papiros/mits.pdf)

**Learning Long-Term Dependencies in NARX Recurrent Neural Networks**

* **http://deeplearning.cs.cmu.edu/pdfs/Narx.pdf**