EpiData Fall 2022 Demo Day Deliverable

datares

Nov. 29

Meet the Team

Aditya Bharath

Alex Chen

Alyssa Liu

Anvesha Dutta

Haoran Jia

Jane Zou

Project Scope

- **Internet of Things:** collect and ingest data across multiple devices, apply predictions on data samples
  - Applications: manufacturing, energy management, building automation

- **Goals:**
  - Develop and deploy AI/ML models on streaming IoT data from two publicly available datasets
  - Rewrite EpiData application using reusable function blocks and publish on GitHub
  - Download and become familiar with the EpiData platform

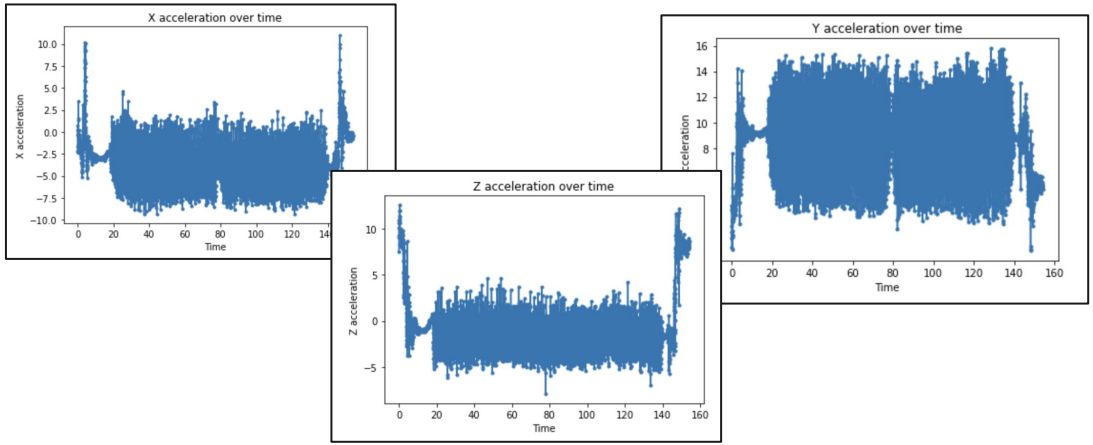
Exploratory Data Analysis

- Regression dataset: 10 minutes for 4.5 months, the **ZigBee wireless sensor network** recording temperature and humidity conditions
  - Kitchen, laundry, living room data importance high
  - Appliances, Visibility, T6, Tdewpoint, Lights, RH_Out, T2

- Classification dataset: **Android smartphone** positioned in chest pocket from 22 participants **walking** over path
  - Time-step, X acceleration, Y acceleration, Z acceleration
  - Y concentrated, Z less variance

LSTM

- Goal: **User Identification** from time series accelerometer data
- Currently employing a Long-Short Term Memory (LSTM) model on training and test data derived from 22 users
- Example of intended model:



LSTM

User 1 User 2 User 3

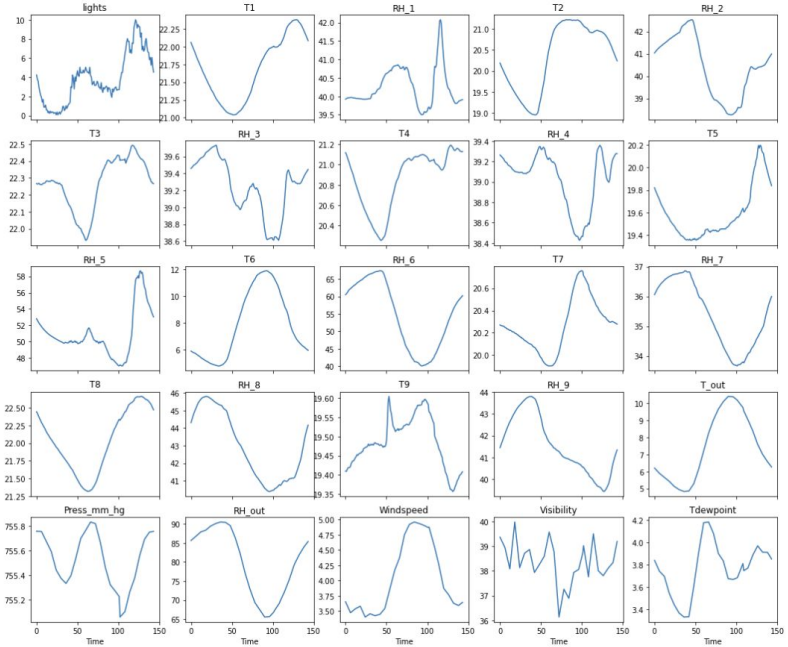
OUTPUT: User Identification

Neural Network

- Goal: Predict energy use (wh) in a house
- Model: 5 - Layer Neural Network Structure
- Evaluation Metric: mean absolute error

Layer (type)	Output Shape	Param #
dense (Dense)	(None, 128)	5120
dense_1 (Dense)	(None, 256)	33024
dense_2 (Dense)	(None, 256)	65792
dense_3 (Dense)	(None, 256)	65792
dense_4 (Dense)	(None, 1)	257

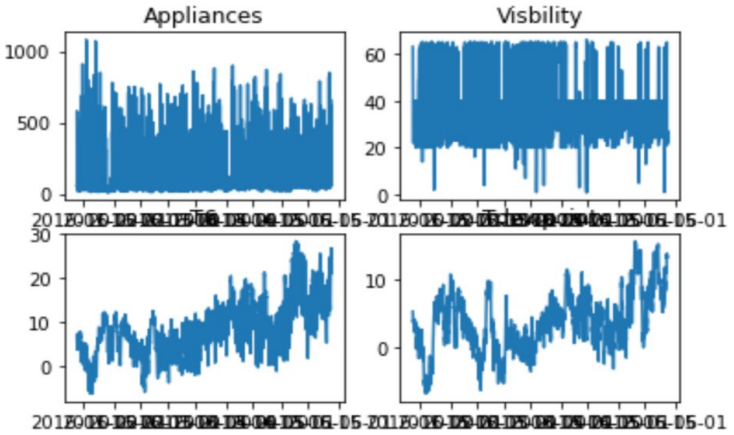
Total params: 169,985
Trainable params: 169,985
Non-trainable params: 0

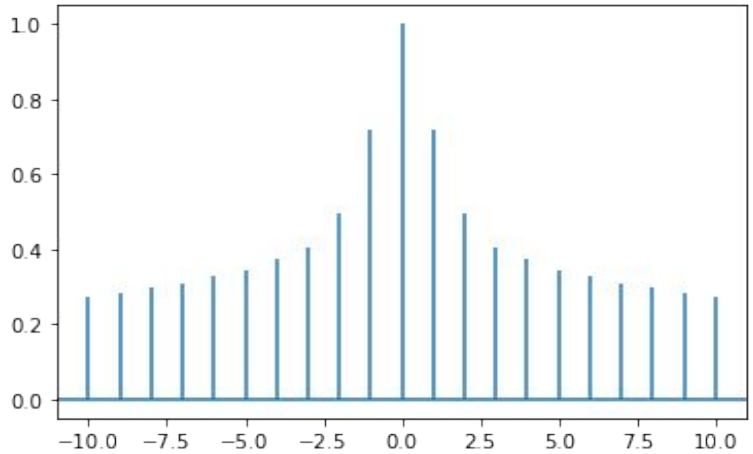


Epoch 500/500
494/494 [==============================] - 1s 1ms/step - loss: 31.0499 - mean_absolute_error: 31.0499 - val_loss: 44.4173 - val_mean_absolute_error: 44.4173

Time Series Regression

- Additive trend in Appliances and Visibility
- Multiplicative trend in T6 and Tdewpoint
- Candidate models : ARIMA, Seasonal ARIMA, One step forward ARIMA (to be decided which one)
- Autocorrelation plots reflect a symmetric correlation pattern, where entries n time periods ahead and behind are highly correlated





Conclusion and Next Steps

- Fall 2022
  - Project Scope and Understanding: 1 week
  - Setup, Compile, Launch Instructions for EpiData Platform: 4 weeks
  - EDA and Modelling: 2 weeks

- Project Continuing **Winter 2023**
  - Modelling for classification and regression
  - Model evaluation and model candidate selection
  - Deployment to GitHub
  - Medium Article