

Minutes Tuesday 1/06 - Gustavo Batista

Lecturer : Gustavo Batista

Date : Tuesday 1/06/21, 7-8pm

Attendees : Gustavo Batista, James Cleaver, Md Ruhul Amin Sarker, Jittinun (Nee) Trairattanasirikul

Apologies : Peter Morian,

Minutes:

- Any suggestions for newer of forecasting models
 - Prophet - Facebook new forecasting model - facebook.github.io/prophet/docs/quick_start.html
 - Univariate model, no need to tune hyperparameters, easy to train
 - Traditional Neural network
- Go through our base model -
https://github.com/Jimbobaus/project---What-Watts---Group-A/blob/main/src/Sandpit_for_Capstone.ipynb
 - Try to use only 2yrs data for training as looking too many years back, might not help with older data
 - Test data - 1 year to cover all the seasons
 - Metrics - look into extreme events where the model not perform well. Largest area,
 - Selling those models to the users - to predict the extreme event e.g. heat wave unexpected, 7 day forecast, graph data diff > 600
 - Time series cross-validation
 - Use 2yrs train, test the next year, use the next 2 yrs to train, etc.
 - Can be very time consuming
- RMSE
 - **Our model is impressive!**
 - Time series needs to be performed consistently especially on those extreme days.
 - Answer about why these features work, etc.
 - Where to get data now for forecasting e.g. PV data, to put value adding in the report
 - E.g. PV data improve by x%, how much and complex to get this data if need to do it right now.
- LSTM - perform better than our existing model?
 - Generally, LSTM is better for large dataset. But no guarantee.
 -
- ARIMA vs SARIMA
 - ARIMA mode uses integrated part (traditional model), not expected to be better than our existing regression
 - SARIMA - seasonal ARIMA model, easier to train
- Ensemble model? - random forecast, support vector machine (SVM), XGBoost (randomforecast with improvement)
- **Expected performance: 1. LSTM 2. XGBOOST 3. Our model 4. SARIMA**
- Don't go to deep in the model, focus on analysis
- One-hot encoding to remove last column of data
- What to expect in the reflective writing
 - Open ended, how interaction in the group, if any conflict or issue
- Follow up on some example of presentation video
 - To follow up with Pierre

Actions:

- Train test split
 - Train data only 2yrs data for training as looking too many years back, might not help with older data
 - Test data - 1 year to cover all the seasons
- Metrics - look into extreme events where the model not perform well
 - Temperature categorical
 - Add 7 days forecast data e.g. heat wave
 - Graph data > 600
- SARIMA - seasonal ARIMA model
- **Expected performance: 1. LSTM 2. XGBOOST 3. Our model 4. SARIMA**

Next meeting:

- Team standup and sync tonight 7:30pm?