Xtreme Weather Forecasting: W3 Team update

GOAL: Review Forcasting Model - Task 2

Name: Teklel/Kelly
Date: Feb 12, 2023

Agenda

- Review research papers
- Identify Relevant Models
- Model Performance Metrics
- Evaluation of models using the dataset

TIMELINE OF THE PROJECT

WE ARE HERE

WEEK 0
JAN 28, 2023

WEEK 1
FEB 4, 2023

WEEK 2
FEB 11, 2023

WEEK 4
FEB 18, 2022

FINAL PRESENTATION
FEb25, 2022

CELEBRATE!!

WEEKLY PLAN

- Semi-self driven project.
- You can go as far as you want.
- Anyone can propose a solution
- Anyone can work on any task they want(don’t re-invent). Leverage.

Week 1	**Week 2**	**Week 3**	**Week 4**
– Data Pre-processing -Data Insights - R**esearch**	– Exploratory Data Analysis (EDA) - Plots -Feature Engineering	-Model Development -Model Evaluation - Initiate deployment	-Submit to Datathon (Optional) - Deployment Build a Dashboard or App

Teams

TASK #	TASK LEADER
TASK 0: Knowledge	Nancy (Final Report)
TASK 1: EDA	Vishu & Pooja
TASK 1a: Feature Engineering	Pranjal
TASK 3: Model development	Tekle & Kelly

Meeting Cadence

Meeting	Day and time
Weekly Team Meeting	Sunday 8 AM CST | 16 UTC | 9:30PM IST
TASK 0: Knowledge	TDB
TASK 1: EDA	Sunday 14 UTC
TASK 1A : FEATURE ENGINEERING	TBD
TASK 4: MODELING	Sunday 17 UTC
Learning & Education Series/Workshops	Every Saturday 16 UTC(based on volunteers)

Task updates

Task 0 - Knowledge

Team Leader:

On-going tasks:

- Maintain Q&A

Last week tasks/results:

- Working with task-1 for data collection and pre-processing.
- …….

Next steps for the upcoming week:

- Work with task-4/6 on ML model related research.
- ML pipelines learning session by Chinmay Kulkarni

Task 1: EDA

Team Leader:

On-going tasks:

- Reviewing the dataset

Last week tasks/results:

- Identified all important columns
- ……

Next steps for the upcoming week:

- creating a data dictionary(containing the meaning of each row in our dataset with help of domain experts)
- ……

Task Reports

Individual Task report

notebooks are available under task-3-eda folder in github

Task 1a – FEATURE ENGINEERING

Team Leader:

On-going tasks:

- Check for outlier estimation, variable reduction(complete any missing data preprocessing tasks)
- Generalising facility_type with respect to median of site_eui and energy consumption both
- Imputation and merging of wind attribute to be done
- Dropping of columns in final Dataset to be done.
- Discussions on more features are going on. More visualisation is being performed to see if data can be more optimised.

Last week tasks/results:

1. Started with checking correlation of features using correlation matrix.
2. Identified Null columns in the data and discussed various imputation methods
3. Identified that there is lot of ones in wind attribute mostly.
4. Identified columns to drop.
5. Grouped the months with respect to season for dimension reduction.
6. Worked on imputing days_with_fog.
7. Various attributes are discussed to asses its importance. For eg, Elevation.

Next steps for the upcoming week:

- Would suggest that we start modelling and then come back to eda now after all missing values are filled.
- This would make it clear what more optimisations we can do to data. What correlation needs to be improved and to be worked on.

Task 3 – Model Development

Team Leader:

On-going tasks:

- Review research papers
- Identify Relevant Models
- Model Performance Metrics
- Evaluation of models using the dataset

Last week tasks/results:

Next steps for the upcoming week:

- Continue with Evaluation of Models

Report:

<Link>