Machine Learning Engineer Nanodegree

Capostone Proposal

Wind Energy Forecasting using Machine-Learning Techniques

Author: José María Dorado Vázquez

Udacity

Sevilla, 2020

Wind Energy Forecasting using Machine-Learning Techniques

Autor:
José María Dorado Vázquez

Udacity

1 PROPOSAL

Wind Energy Forecasting using data from a wind turbine (Kaggle dataset) [1].

2 DOMAIN BACKGROUD

Wind power forecast evaluation is of key importance for forecast provider selection, forecast quality control, and model development. While forecasts are most often evaluated based on squared or absolute errors, these error measures do not always adequately reflect the loss functions and true expectations of the forecast user, neitherdo they provide enough information for the desired evaluation task. Over the last decade, research in forecast verification has intensified, and a number of verification frameworks and diagnostic tools have been proposed. However, the corresponding literature is generally very technical and most often dedicated to forecast model developers. This can make forecast users struggle to select the most appropriate verification tools for their application while not fully appraising subtleties related to their application and interpretation.

Wind energy is of vital importance among the low-carbon energy supply and constitutes a keystone component for micro-grids in a way towards the smart grid infrastructure. However, stochastic and intermittent wind power generation poses a number of challenges to the large scale penetration of wind power. These wind-related uncertainties can put the system reliability and power quality at risk with the increasing penetration of wind power and thus, the main grid integration issues such as balance management and reserve capacities can come into question . Reducing the need for balancing energy and making the power generation scheduling and dispatch decisions can be realised with the help of wind speed and power generation forecasts..Furthermore,the forecasts can play a pivotal role in keeping the costs competitive by reducing the need for wind curtailments and thereby, increasing revenue in electricity market operations. However, the random and unstable characteristics of the wind make it considerably difficult to forecast the wind speed and power accurately.

In the literature, many forecasting approaches have been studied and proposed, each utilising a different technique and performing well with a different prediction horizon. Recent studies in the area of wind prediction are predominantly focused on the short term wind predictions ranging from minutes to a few days due to the importance of these data on power systems. Especially day-ahead predictions are of significant interest for system operations such as scheduling, unit commitment and load following. However, it is generally difficult to accomplish such a long-term prediction and moreover, the approaches designed for long prediction horizons may be deficient for shorter terms in terms of prediction performance. Following many studies in the wind forecasting field, it can be indicated that, to date, the targeted performance levels have not been attained with the individual models due to the fact that these models cannot give satisfactory results for all situations. For instance, the physical models produce coarse predictions for short-term horizons while mostly outperform the other models in medium- and long-term horizons. Also, Artificial Intelligent (AI) based models that rely on a large number of historical data for constructing an input/output mapping function can be less effective than some basic conventional statistical methods for certain application areas in the case of inadequate available information [2].

3 PROBLEM STATEMENT

The goal of this project is to predict the energy generation from a win-turbine using meteorological variables.

To achieve this goal, it will be necessary a feature extraction and evaluation of the data understand the behaviour of a wind turbine. Then, several models will be realised, since linear models to complex deep-learning structures using Tensorflow.

It will be interesting to develop an approach using time-series and statistical approaches to predict the wind-energy forecasting 3-hours-ahead. It is proposed created an ARIMA model using historical wind power generation.

In summary, the main objectives are as follows:

- Develop different statistical and AI models to predict energy forecasting using meteorological data. Some of these models could be:
  - Linear Regression
  - Multilinear Regression
  - XGBoost
  - Regression Forest
  - SVR
  - Deep Neural Network
- Optional: Predict the wind energy 3-hour-ahead using:
  - ARIMA
  - LSTM
  - CNN-LSTM

4 DATASET AND INPUTS

Project : SCADA Data of a Wind Turbine in Turkey¶

Context. In Wind Turbines, Scada Systems measure and save data's like wind speed, wind direction, generated power etc. for 10 minutes intervals. This file was taken from a wind turbine's SCADA system that is working and generating power in Turkey.

- **Content** The data's in the file are:
- **Date/Time** (for 10 minutes intervals)
- **LV ActivePower (kW)**: The power generated by the turbine for that moment
- **Wind Speed (m/s)**: The wind speed at the hub height of the turbine (the wind speed that turbine use for electricity generation)
- **TheoreticalPowerCurve (KWh)**: The theoretical power values that the turbine generates with that wind speed which is given by the turbine manufacturer
- **Wind Direction (°)**: The wind direction at the hub height of the turbine (wind turbines turn to this direction automatically)

Details of the dataset:

- Number of rows: 50.530 rows
- Number of columns: 6 columns
    - DateTime 50.530 non-null. Type : Object
    - LVA_Power: 50.530 non-null. Type : Float 64
    - WindSpeed: 50.530 non-null. Type : Float 64
    - TPC_KWh: 50.530 non-null. Type : Float 64
    - WindDirection: 50.530 non-null. Type : Float 64

5 SOLUTION STATEMENT

The proposed solution to this problem is to apply different Machine-Learnings and Deep-Learning algorithms that have proved to be highly successful in the field of regression problems.

We need to analyse the data to understand the relationship between inputs and output. First, it will be necessary to search for outliers and NaNs values that can be a problem at an hour of train hour models. Also, some particular and extreme values it will be analysed in order to these values will be physically correct.

The next step will be to develop a Deep Neural Network using Tensorflow and compared the results with another methods.

We will used the typical error metrics on regression problems:

- R2 Score
- Mean Squared Error

6 BENCHMARK MODEL

One of the choices of this dataset that few people have dealt with, allowing it to be one of the first solutions proposed in Kaggle. Most of the people who have proposed solutions have been involved in the feature extraction part.

The results found are as follows:

- XGBRegressor: R2_score in testing set: 0.8823
- LGBMRegressor: R2_score in testing set: 0.9108

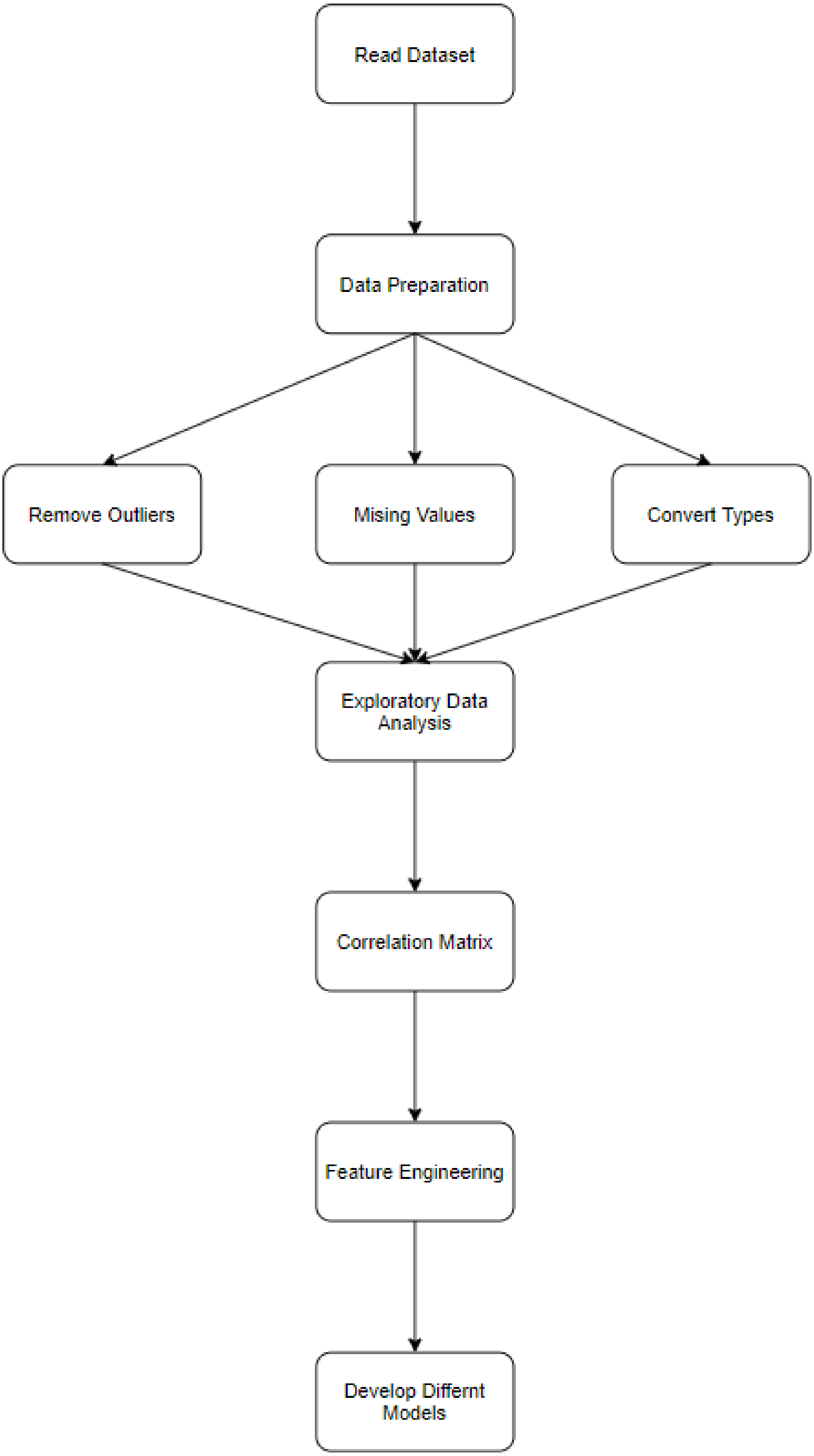
7 EVALUATION METRICS

It will be used the classical evaluation metrics in a regression problem:

- R2 score
- Mean Squared Error

8 PROJECT DESSING

The proposed flowchart of the different steps to complete the project is as follow:


Read Dataset
Data Preparation
Remove Outliers
Mising Values
Convert Types
Exploratory Data Analysis
Correlation Matrix
Feature Engineering
Develop Differnt Models

9 References

[1] B. Erisen, “Kaggle,” 2018. [Online]. Available: https://www.kaggle.com/johndev001/wind-data-analysis-with-some-feature-engineering. [Accessed 2020 May 24].

[2] A. Tascikaraoglu and M. Uzunoglu, “A review of combined approaches for prediction of short-term wind speed and power,” *Renewable and Sustainable Energy Reviews,* vol. 4, pp. 243-254, 2014.